Selecting optimal block lengths for block bootstrap methods

S.N. Lahiri
Department of Statistics
Iowa State University
Ames, IA 50011
email:snlahiri@iastate.edu

ABSTRACT

In this article, we review some general methods for data-based selection of the optimal block lengths for block bootstrap methods. We also report the results of a small simulation study on finite sample performance of the methods.

Key words and Phrases: Jackknife-after-bootstrap method, mean squared error, plug-in method, subsampling method, stationarity.

1 Introduction

Bootstrapping time series data often involves the use of blocks. As is well known from various studies on the problem (cf. Künsch (1989) and Hall, Horowitz and Jing (1995)), accuracy of block bootstrap estimators critically depends on the block size. In this paper, we review and compare some data-based methods for selecting the optimal block lengths for block bootstrap methods. There are essentially two approaches to the selection of the optimal block size, namely (i) plug-in methods and (ii) empirical criteria-based methods. For the plug-in methods, one needs to have available an expression for (the leading term(s) of) the theoretical optimal block size. The population characteristics appearing in this expression are then estimated and 'plugged' into the given expression to produce the plug-in estimator of the optimal block size. A major disadvantage of this approach is that one has to invest a huge amount of effort to derive the explicit expression for the theoretical optimal block size analytically, which often depends on various population parameters in an intricate manner. Furthermore, the plug-in estimators are target specific in the sense that altogether different analytical considerations may be needed for different target functionals of interest, as one must derive the expressions for the corresponding optimal block sizes. The second approach is typically more widely applicable, and bypasses much of the analytical considerations required for deriving explicit forms of the theoretical optimal block sizes. Here one defines a general method for producing an estimator of a criterion function (such as the mean squared error) and minimizes the estimated criterion function to get an estimator of the optimal block size. Such empirical methods are often applicable to various target functionals of interest, without requiring extensive analytical work. The price paid for such wider applicability of these methods is a more demanding level of computation.

In a seminal paper, Hall, Horowitz, and Jing (1995) developed an estimated criterion-based method for choosing the MSE-optimal block sizes in various inference problems. They employ the subsampling method of Politis and Romano (1994) to construct an estimate of the MSE-function. In the same vein, one may employ other resampling methods, including (a second level of) block bootstrapping itself, to generate alternative estimates of the MSE function (cf. Chapter 7, Lahiri (2003)). In a recent work, a hybrid method, called the nonparametric plug-in method has been proposed by Lahiri, Furukawa, and Lee (2003) which combines the desirable features of both plug-in methods and criteria-based methods. The nonparametric plug-in method employs a suitable resampling method, known as the Jackknife-After-Bootstrap (or the JAB) method of Efron (1992) and Lahiri (2002), to construct estimates of the leading terms of the optimal block size in an implicit way. Indeed, the nonparametric plug-in method neither requires explicit expressions for the population characteristics in the theoretical optimal block size nor does it involve minimization of an estimated criterion function. Furthermore, the nonparametric plug-in method can be applied to different inference problems like the Hall, Horowitz, and Jing (1995) method and hence, is not a target specific method as standard plug-in methods. In the next few sections, we describe the Hall, Horowitz, and Jing (1995) method and the nonparametric plug-in method and compare their finite sample performance through a small numerical example.

The rest of the paper is organized as follows. We conclude Section 1 with a brief literature review. In Section 2, we describe some resampling methods that are relevant for describing the Hall, Horowitz, and Jing (1995) and the nonparametric plug-in methods. In Section 3, we describe the Hall, Horowitz, and Jing (1995) and the nonparametric plug-in methods. In section 4, we report the results of a small simulation study and in Section 5, we make some concluding remarks.

Different versions of the block bootstrap method for dependent data have been put forward by Hall (1985), Carlstein (1986), Künsch (1989), Liu and Singh (1992), Politis and Romano (1992), Carlstein et al. (1998), and Paparoditis and Politis (2001), among others. For results on properties of block-bootstrap methods, see Lahiri (1991, 1996, 1999), Bühlmann (1994), Naik-Nimbalkar and Rajarshi (1994), Hall, Horowitz and Jing (1995), Götze and Künsch (1996), and the references therein. Plug-in methods for choosing the optimal block size have been developed by by Bühlmann and Künsch (1999) and Politis and White (2003), among others. A book length treatment of various issues involving resampling methods for dependent data is given in Lahiri (2003).

2 Preliminaries

2.1 The Moving Block Bootstrap Method

For brevity, we shall restrict our attention to a commonly used block bootstrap method, known as the Moving Block Bootstrap (MBB) method, independently proposed by Künsch (1989) and Liu and Singh (1992). Let $\{X_1, \ldots, X_n\} \equiv \mathcal{X}_n$ denote the observations from a stationary time series $\{X_i\}_{i \in \mathbb{Z}}$ and let

$$T_n = t_n(\mathcal{X}_n; \theta) \tag{2.1}$$

be a random variable of interest, where θ is a population parameter. For example, we may have

$$T_n = \sqrt{n}(\bar{X}_n - \mu), \tag{2.2}$$

where $\bar{X}_n = n^{-1} \sum_{i=1}^{n} X_i$ denotes the sample mean and $\mu = EX_1$ denotes the population mean. Here our objective is to estimate the sampling distribution of T_n and its functionals. Let $\ell = \ell_n$ be an integer such that

$$\ell \to \infty \quad \text{and} \quad n^{-1}\ell \to 0, \tag{2.3}$$

as $n \to \infty$. That is, the block size ℓ tends to infinity but at a rate slower than the sample size n. Let $N = n - \ell + 1$ and let $\mathcal{B}_i = (X_i, \ldots, X_{i+\ell-1})'$, $i = 1, \ldots, N$ denote the collection of all overlapping blocks of length ℓ contained in $\mathcal{X}_n$. Also, let $b_1 = \lceil n/\ell \rceil$, where for any real number x, $\lceil x \rceil$ denotes the smallest integer not less than x. The MBB selects b_1 blocks at random, with replacement from $\{\mathcal{B}_1, \ldots, \mathcal{B}_N\}$ to generate $n_1 \equiv b_1 \ell$ many bootstrap observations, obtained by concatenating the elements from the b_1 resampled blocks. Let $\{X_1^*, \ldots, X_n^*\} = \mathcal{X}_n^*$ denote the first n block bootstrap observations. Then, the MBB version of T_n is given by

$$T_n^* = t_n(\mathcal{X}_n^*; \hat{\theta}_n), \tag{2.4}$$

where $\hat{\theta}_n$ is a suitable estimator of the parameter θ based on $\mathcal{X}_n$. For example, for $T_n = \sqrt{n}(\bar{X}_n - \mu)$, its bootstrap version is given by $T_n^* = \sqrt{n}(\bar{X}_n^* - \hat{\mu}_n)$, where $\bar{X}_n^* = n^{-1} \sum_{i=1}^{n} X_i^*$ is the bootstrap sample mean and where $\hat{\mu}_n = E_*(\bar{X}_n^*)$ is the conditional expectation of $\bar{X}_n^*$, given $\mathcal{X}_n$. Let G_n denote the distribution of T_n. The MBB estimator $\hat{G}_n$ of G_n is given by the conditional distribution of T_n^* given $\mathcal{X}_n$:

$$\hat{G}_n(x) = P_*(T_n^* \le x), x \in \mathbb{R}, \tag{2.5}$$

where P_* denotes the conditional probability given $\mathcal{X}_n$. For a given functional ψ, the MBB estimator of a population characteristic $\psi_n = \psi(G_n)$ is given by

$$\hat{\psi}_n \equiv \hat{\psi}_n(\ell) = \psi(\hat{G}_n). \tag{2.6}$$

For example, if $\psi_n = \mathrm{Var}(T_n)$, which corresponds to the variance functional $\psi(G) \equiv \int x^2 dG(x) - (\int x dG(x))^2$, the MBB estimator of ψ_n is given by $\hat{\psi}_n(\ell) = \psi(\hat{G}_n) = \mathrm{Var}_*(T_n^*)$, the conditional variance of T_n^* given $\mathcal{X}_n$.

2.2 The Jackknife-After-Bootstrap variance estimator

The JAB method was initially proposed by Efron (1992) to assess accuracy of bootstrap estimators for independent data. An extension of the method to the dependent case has been given by Lahiri (2002). As in the case of independent data, the JAB method of Lahiri (2002) produces a variance estimate by merely regrouping the resampled blocks, generated for the computation of the original block bootstrap estimator and does not require any additional resampling. The key construction that ensures this desirable feature of the JAB uses a 'blocks of blocks' version of the block jackknife method. To describe the JAB method, suppose that $\hat{\psi}_n(\ell)$ is the MBB estimator of ψ_n based on blocks of size ℓ, where ψ_n is a population characteristic of the distribution of T_n of (2.1), such as, $\psi_n = \mathrm{Var}(T_n)$. Let $m \equiv m_n$ be an integer such that as $n \to \infty$,

$$m \to \infty \quad \text{and} \quad m/n \to 0. \tag{2.7}$$

Here, m denotes the number of bootstrap blocks to be deleted. With $N = n - \ell + 1$, let $M = N - m + 1$ and for $i = 1, \ldots, M$, let $I_i = \{1, \ldots, N\} \setminus \{i, \ldots, i + m - 1\}$ denote the index set of all blocks of size ℓ obtained by deleting the m blocks $\{\mathcal{B}_i, \ldots, \mathcal{B}_{i+m-1}\}$. Then, to generate the ith block-deleted jackknife point value $\hat{\psi}_n^{(i)}$, we resample b_1 blocks randomly, with replacement from the collection $\{\mathcal{B}_j : j \in I_i\}$, compute the corresponding block bootstrap estimator $\hat{G}_{n,i}$ (say) of G_n, and then, apply the functional ψ to $\hat{G}_{n,i}$, to get $\hat{\psi}_n^{(i)} = \psi(\hat{G}_{n,i})$, $i = 1, \ldots, M$. The JAB estimator of the variance of $\hat{\psi}_n$ is given by

$$\widehat{\mathrm{VAR}}_{\mathrm{JAB}}(\hat{\psi}_n) = \frac{m}{(N-m)} \frac{1}{M} \sum_{i=1}^{M} (\tilde{\psi}_n^{(i)} - \hat{\psi}_n)^2 \tag{2.8}$$

where $\tilde{\psi}_n^{(i)} = m^{-1}[N\hat{\psi}_n - (N-m)\hat{\psi}_n^{(i)}]$ is the ith *block-deleted jackknife pseudo-value* of $\hat{\psi}_n$, $i = 1, \ldots, M$.

Monte-Carlo computation of the jackknife point values $\hat{\psi}_n^{(i)}$ proceeds as follows. Suppose that $\{{}_k\mathcal{B}_1^*, \ldots, {}_k\mathcal{B}_{b_1}^*\}, k = 1, \ldots, K$ denote the K conditionally iid sets of bootstrap blocks drawn randomly and with replacement from the full collection $\{\mathcal{B}_1, \ldots, \mathcal{B}_N\}$ of overlapping blocks, for Monte-Carlo evaluation of the given block bootstrap estimator $\hat{\psi}_n$. For $1 \leq i \leq M$, let

$$I_i^* = \Big\{k : 1 \leq k \leq K, \{{}_k\mathcal{B}_1^*, \ldots, {}_k\mathcal{B}_{b_1}^*\} \cap \{\mathcal{B}_i, \ldots, \mathcal{B}_{i+m-1}\} = \emptyset\Big\}. \tag{2.9}$$

Thus, I_i^* denotes the subcollection of all resampled block-sets $({}_k\mathcal{B}_1^*, \ldots, {}_k\mathcal{B}_{b_1}^*)$ of size b_1 where none of the resampled blocks ${}_k\mathcal{B}_1^*, \ldots, {}_k\mathcal{B}_{b_1}^*$ equals $\mathcal{B}_i, \ldots, \mathcal{B}_{i+m-1}$. Then each such $\{{}_k\mathcal{B}_1^*, \ldots, {}_k\mathcal{B}_b^*\}$ represents a random sample from the 'blocks of blocks'-deleted collection $\{\mathcal{B}_j : j \in I_i\}$ (cf. Lahiri (2002)). Let $T_n^{*(i)}$ denote the bootstrap version of T_n based on the 'delete-m' collection of blocks $\{\mathcal{B}_j : j \in I_i\}$. Then, $\hat{G}_{n,i}(\cdot) = P_*(T_n^{*(i)} \leq \cdot)$ and the i-th jackknife point value $\hat{\psi}_n^{(i)}$ is given by $\hat{\psi}_n^{(i)} = \psi(\hat{G}_{n,i})$. For Monte-Carlo evaluation of $\hat{\psi}_n^{(i)}$, we *extract* the subcollection $\{({}_k\mathcal{B}_1^*, \ldots, {}_k\mathcal{B}_b^*) : k \in I_i^*\}$ from $\{({}_k\mathcal{B}_1^*, \ldots, {}_k\mathcal{B}_b^*) : 1 \leq k \leq K\}$ and *reuse* the corresponding bootstrap observations to compute the bootstrap replicates of $T_n^{*(i)}$. As a result, the replicates of $T_n^{*(i)}$ may be computed without any additional resampling.

Next, denote the empirical distribution of the replicates of $T_n^{*(i)}$ by $G_{n,i}^{\#}$. Then, $G_{n,i}^{\#}$ is the Monte-Carlo approximation to $\hat{G}_{n,i}$ and the Monte-Carlo approximation to the i-th jackknife point value $\hat{\psi}_n^{(i)}$ is given by $\psi_n^{(i)\#} = \psi(G_{n,i}^{\#})$. For example, for the variance functional $\psi_n = \mathrm{Var}(T_n)$, the Monte-Carlo approximation $\psi_n^{(i)\#}$ is given by the *sample* variance of the replicates of $T_n^{*(i)}$, generated as described above. For further details and discussion of the properties of the JAB for dependent data, see Lahiri (2002, 2003).

2.3 The Subsampling Method

Let T_n, given by (2.1) be a random variable of interest and let m be an integer satisfying (2.6). Like the MBB, the objective of the subsampling method is to obtain estimates of the sampling distribution G_n of T_n and its functionals. Let $\mathcal{B}(i;m) = (X_i, \ldots, X_{i+m-1})$, $i = 1, \ldots, n-m+1$ denote the overlapping subsamples of size m. On each subsample $\mathcal{B}(i;m)$, define a subsample copy $T_{i,m}$ of T_n by replacing $\mathcal{X}_n$ with $\mathcal{B}(i;m)$, n with m, and θ with a suitable estimator $\hat{\theta}_n$ as:

$$T_{i,m} \equiv t_m(\mathcal{B}(i;m); \hat{\theta}_n). \tag{2.10}$$

Here the estimator $\hat{\theta}_n$ is defined using all n observations. For example, for $T_n = n^{1/2}(\bar{X}_n - \mu)$, we may set $T_{i,m} = n^{1/2}(\bar{X}_{i,m} - \bar{X}_n)$, $i = 1, \ldots, n-m+1$, where $\bar{X}_{i,m} = m^{-1}\sum_{k=i}^{i+m-1} X_k$ is the average of the m elements in the subsample $\mathcal{B}(i;m)$ and where the population mean μ is estimated by the sample mean $\bar{X}_n$. Another frequently used alternative estimator of μ is given by

$$\hat{\mu}_n = (n-m+1)^{-1} \sum_{i=1}^{n-m+1} \bar{X}_{i,m},$$

the average of the subsample averages.

The subsample estimator of G_n based on overlapping subsamples of size m is given by the empirical distribution of $T_{i,m}$, $i = 1, \ldots, n-m+1$, i.e., by

$$\check{G}_{n,m}(x) = (n-m+1)^{-1} \sum_{i=1}^{n-m+1} \mathbb{1}(T_{i,m} \le x), \quad x \in \mathbb{R}, \tag{2.11}$$

where $\mathbb{1}(\cdot)$ denotes the indicator function, with $\mathbb{1}(S) = 1$ or 0 according as the statement S is true or false. Likewise, the subsampling estimator of a functional $\psi_n = \psi(G_n)$ is given by $\check{\psi}_n \equiv \check{\psi}_{n,m} = \psi(\check{G}_{n,m})$. Thus, for T_n of (2.2), if we write $\bar{T}_n$ for the average of the subsampling copies $\{T_{i,m} : i = 1, \ldots, n-m+1\}$, then the subsampling estimator of $\sigma_n^2 \equiv \mathrm{Var}(T_n)$ is given by

$$\begin{aligned} \check{\sigma}_n^2 &= \frac{1}{n-m+1} \sum_{i=1}^{n-m+1} \Big(T_{i,m} - \bar{T}_n\Big)^2 \\ &= \frac{1}{n-m+1} \sum_{i=1}^{n-m+1} \Big(\bar{X}_{i,m} - \hat{\mu}_n\Big)^2. \end{aligned}$$

For more details on the properties and applications of the subsampling method, see Politis, et al. (1999).

3 Methods for selecting optimal block sizes

3.1 The basic framework

Let ψ_n be a target parameter and let $\hat{\psi}_n(\ell)$ be its MBB estimator based on blocks of length ℓ and based on the sample $\mathcal{X}_n$. We shall suppose that ψ_n and $\hat{\psi}_n(\ell)$ are such that the following expansion for the bias and variance of $\hat{\psi}_n(\ell)$ over a suitable set $\mathcal{L}_n \subset \{2, \ldots, n\}$ of possible block lengths ℓ hold:

$$n^{2a} \cdot \mathrm{Var}(\hat{\psi}_n(\ell)) = C_1 n^{-1} \ell^r + o(n^{-1}\ell^r) \ \text{ as } n \to \infty \tag{3.1}$$

$$n^a \cdot \mathrm{Bias}(\hat{\psi}_n(\ell)) = C_2 \ell^{-1} + o(\ell^{-1}) \ \text{ as } n \to \infty, \tag{3.2}$$

where C_1, C_2 are population parameters, $r \geq 1$ is an integer, and $a \in [0, \infty)$ is a known constant. For example, for $\psi_n = n\mathrm{Var}(\hat{\theta}_n)$ and $\hat{\theta}_n =$ the sample mean, (2.1) and (2.2) hold with $r = 1$, $a = 0$ (cf. Hall, Horowitz, and Jing (1995)), while for bootstrap estimation of the distribution function ψ_n of studentized $\hat{\theta}_n$, (2.1) and (2.2) hold with $r = 2$ and $a = 1/2$ (cf. Lahiri (2003)). In these examples, we may take $\mathcal{L}_n = [n^\epsilon, n^{\frac{1}{2}-\epsilon}]$ for some $\epsilon \in (0, 1/2)$ for the variance functional and we may take $\mathcal{L}_n = [n^\epsilon, n^{\frac{1}{3}}]$ for some $\epsilon \in (0, 1/3)$ for the distribution function. Exact expressions of the parameters C_1 and C_2 are typically very complicated, involving cumulant spectral density of the process $\{X_i\}$ of order two and higher. However, unlike traditional plug-in methods, this does not pose a problem, as explicit expressions for these quantities are not needed for applying either the Hall, Horowitz, and Jing (1995) method or the nonparametric plug-in method described below.

Note that (3.1) and (3.2) readily yield an expansion for the MSE of $\hat{\psi}_n(\ell)$:

$$n^{2a} \cdot \mathrm{MSE}(\hat{\psi}_n(\ell)) = C_1 n^{-1} \ell^r + C_2^2 \ell^{-2} + o(n^{-1}\ell^r + \ell^{-2}) \ \text{ as } n \to \infty. \tag{3.3}$$

Hence, the MSE-optimal block size, say, ℓ_n^0 has the form

$$\ell_n^0 \sim \left[\frac{2C_2^2}{rC_1}\right]^{\frac{1}{r+2}} n^{\frac{1}{r+2}}, \tag{3.4}$$

where for any two sequences of positive real numbers $\{s_n\}_{n\geq 1}$ and $\{t_n\}_{n\geq 1}$, we write $s_n \sim t_n$ if $\lim_{n\to\infty} s_n/t_n = 1$. In the next two sections, we describe the Hall, Horowitz, and Jing (1995) method and the nonparametric plug-in method for estimating the optimal block length parameter ℓ_n^0.

3.2 The Hall, Horowitz and Jing method

Let m be an integer satisfying (2.7). Given observations $\{X_1, \ldots, X_n\}$, consider the maximal overlapping subsamples of $\{X_1, \ldots, X_n\}$ of size m, given by $\mathcal{B}(i, m) = (X_i, \ldots, X_{i+m-1})$, $i = 1, \ldots, n - m + 1$. Next, let ℓ be an integer such that

$$\ell^{-1} + m^{-1}\ell = o(1)$$

as $n \to \infty$. For each $i = 1, \ldots, n - m + 1$, now treat $\mathcal{B}(i, m)$ as the 'full' sample and compute the MBB estimator, $\hat{\psi}_{i,\ell}$, say, of a given parameter ψ_n based on blocks of size ℓ using the m values in the ith subsample $\{X_i, \ldots, X_{i+\ell-1}\}$. Next, using the

MBB estimators $\hat{\psi}_{i,\ell}$'s from the $(n-m+1)$-many subsamples, define an estimator of $\mathrm{MSE}_m(\ell) \equiv \mathrm{MSE}(\hat{\psi}_m(\ell)$, the MSE function corresponding to a sample size m, as:

$$\widehat{\mathrm{MSE}}_m(\ell) = (n-m+1)^{-1} \sum_{i=1}^{n-m+1} (\hat{\psi}_{i,\ell} - \hat{\psi}_n(\ell^*))^2,$$

where $\hat{\psi}_n(\ell^*)$ is the MBB estimator of ψ_n based on a pilot block length ℓ^* and the full sample $\{X_1, \ldots, X_n\}$. In the next step, $\widehat{\mathrm{MSE}}_m(\ell)$ is minimized over a set of ℓ-values to generate an estimator $\check{\ell}_m$, say, of the optimal block size ℓ_m^0 (corresponding to a sample size m). For a theoretical optimal block size ℓ_n^0 that is of the order $n^{\frac{1}{r+2}}$ (cf. (3.4)), the Hall, Horowitz, and Jing (1995) estimator $\bar{\ell}_n$ of the optimal block size ℓ_n^0 (for the sample size n) is then obtained by scaling the estimator $\check{\ell}_m$ as

$$\bar{\ell}_n^0 = [n/m]^{\frac{1}{r+2}} \check{\ell}_m. \tag{3.5}$$

Note that the Hall, Horowitz, and Jing (1995) method requires the user to choose two smoothing parameters, namely, the pilot block length ℓ^* and the subsample size m. To reduce the effect of the initial block length ℓ^* on the final estimate $\bar{\ell}_n^0$, Hall, Horowitz, and Jing (1995) suggested iterating the procedure where, in the second round, ℓ^* is replaced by $\bar{\ell}_n^0$ to generate an updated estimate $\bar{\ell}_n^{0,1}$, say, and the process is repeated with $\ell^* = \bar{\ell}_n^{0,1}$ in the next step, and so on. We consider finite sample convergence properties of the Hall, Horowitz, and Jing (1995) method in Section 4.

3.3 The nonparametric plug-in method

From (3.1) and (3.2), note that

$$C_1 \sim [n\ell^{-r}] \{n^{2a} \cdot \mathrm{Var}(\hat{\psi}_n(\ell))\}$$

and

$$C_2 \sim [\ell] \{n^a \cdot \mathrm{Bias}(\hat{\psi}_n(\ell))\}.$$

Suppose that the population quantities $\mathrm{Var}(\hat{\psi}_n(\ell))$ and $\mathrm{Bias}(\hat{\psi}_n(\ell))$ could be estimated consistently, say, by $\widehat{\mathrm{VAR}}$ and $\widehat{\mathrm{BIAS}}$, respectively, for some sequence $\{\ell_1\}$ satisfying (2.3) in the following sense:

$$\frac{\widehat{\mathrm{VAR}}}{\mathrm{Var}(\hat{\psi}_n(\ell_1))} \to_p 1 \quad \text{as} \quad n \to \infty$$

$$\frac{\widehat{\mathrm{BIAS}}}{\mathrm{Bias}(\hat{\psi}_n(\ell_1))} \to_p 1 \quad \text{as} \quad n \to \infty.$$

Then we may estimate the parameters C_1 and C_2 consistently by the "natural" estimators

$$\hat{C}_1 = [n\ell_1^{-r}] \{n^{2a} \cdot \widehat{\mathrm{VAR}}\}, \tag{3.6}$$

$$\hat{C}_2 = [\ell_1] \{n^a \cdot \widehat{\mathrm{BIAS}}\}. \tag{3.7}$$

The nonparametric plug-in estimator of the optimal block length ℓ_n^0 is given by

$$\hat{\ell}_n^0 = \left[\frac{2\hat{C}_2^2}{r\hat{C}_1}\right]^{\frac{1}{r+2}} n^{\frac{1}{r+2}}. \tag{3.8}$$

Lahiri, Lee, and Furukawa (2003) prescribe the JAB variance estimator of (2.8) (with $\ell = \ell_1$) for $\widehat{\text{VAR}}$ and propose

$$\widehat{\text{BIAS}} \equiv 2(\hat{\psi}_n(\ell_1) - \hat{\psi}_n(2\ell_1)) \tag{3.9}$$

as an estimator of $\text{Bias}(\hat{\psi}_n(\ell_1))$. They also provide conditions on the process $\{X_i\}$ and on the smoothing parameters ℓ_1 and m such that the optimal block length estimator $\hat{\ell}_n^0$ is consistent for the theoretical optimal block length ℓ_n^0 for various target functionals ψ_n, including the variance functional and the quantile functional. For more details on the theoretical properties of the nonparametric plug-in method, see Lahiri, Furukawa, and Lee (2003) and Lahiri (2003). In the next section, we consider finite sample properties of the nonparametric plug-in method in a small numerical study.

4 Simulation Study

In this section, we report the results of a simulation study on finite sample performance of the Hall, Horowitz, and Jing (1995) method and the nonparametric plug-in method. We consider the time series model, given by

$$X_i = Y_i + .2Y_{i-1} + .6Y_{i-2} + 8.0Y_{i-3} \tag{4.1}$$

where $\{Y_i\}$ is a sequence of independent and identically distributed (iid) N(0,1) random variables. We consider bootstrap estimation of the following parameters related to the distributions of $\bar{X}_n$ and its studentized version T_{1n}:

$$\begin{aligned}
\psi_{1n} &= n\text{Var}(\bar{X}_n) \\
\psi_{2n} &= P(T_{1n} \leq 1.0) \\
\psi_{3n} &= 0.20 \text{ quantile of } T_{1n}, \quad \text{and} \\
\psi_{4n} &= 0.70 \text{ quantile of } T_{1n},
\end{aligned} \tag{4.2}$$

where, with $\hat{\tau}_n = |\hat{\gamma}_n(0) + 2\sum_{k=1}^{\ell_1} \hat{\gamma}_n(k)|^{1/2} + n^{-1}$, and $\hat{\gamma}_n(k) = n^{-1}\sum_{i=1}^{n-k}(X_i - \bar{X}_n)(X_{i+k} - \bar{X}_n)$, $0 \leq k \leq n-1$, the studentized statistic T_{1n} is given by

$$T_{1n} = \sqrt{n}(\hat{\theta}_n - \theta)/\hat{\tau}_n. \tag{4.3}$$

Let $\hat{\psi}_{kn}(\ell)$ denote the MBB estimator of ψ_{kn} based on blocks of length ℓ, $k = 1, \ldots, 4$. From Lahiri, Furukawa, and Lee (2003), it follows that under appropriate regularity conditions, expansions (3.1) and (3.2) hold for the variance functional $\hat{\psi}_{1n}(\ell)$ with $a = 0$ and $r = 1$, while (3.1) and (3.2) hold for the other three functionals with $a = 1/2$ and $r = 2$. We define the Hall, Horowitz, and Jing (1995) and the nonparametric plug-in estimators of the optimal block lengths respectively by (3.5) and (3.8) with the above values of a and r for the functionals ψ_{kn}'s. For the simulation study, we set the sample size at $n = 80$.

For each of the four functionals $\psi_{1n} - \psi_{4n}$, the exact expressions of the corresponding constants C_1 and C_2 are known (cf. Lahiri (2003)). Indeed, the constants depend upon some intricate population characteristics of the underlying process, such as the second and third order cumulant spectral density functions of $\{X_i\}$

and their derivatives at zero. As a result, direct estimation of these population parameters is a difficult problem and requires one to decide on specific estimation method for each individual population characteristic and also, choose appropriate smoothing parameters for estimating these characteristics. However, both the Hall, Horowitz, and Jing (1995) and the nonparametric plug-in methods can be applied to all four functionals without the knowledge of the exact expressions for C_1's and C_2's.

For implementing the Hall, Horowitz, and Jing (1995) method, we take $\ell^* = 5$ and $m = 20$, which are comparable to the values suggested by Hall, Horowitz, and Jing (1995) in their simulation study. For implementing the nonparametric plug-in method, Lahiri, Furukawa, and Lee (2003) suggest choosing the initial block size ℓ_1 and the JAB blocking parameter m using the formulas

$$\ell = c_1 \cdot n^{\frac{1}{r+4}} \quad \text{and} \quad m = c_2 \cdot n^{1/3}\ell^{2/3}, \tag{4.4}$$

where $r = 1$ for the variance parameter ψ_{1n} and $r = 2$ for the remaining parameters ψ_{kn}, $k = 2, 3, 4$. On the basis of a numerical study involving similar time series models, Lahiri, Furukawa, and Lee (2003) suggest taking $c_1 = 1.0$ for the initial block length ℓ_1 in (4.4) for all four functionals, which we use here. And for the JAB parameter m in (4.4), we took $c_2 = 1.0$ for the parameter ψ_{1n} and $c_2 = 0.1$ for ψ_{2n}, ψ_{3n}, and ψ_{4n}, also as suggested in Lahiri, Furukawa, and Lee (2003).

Computations for finding the true optimal block sizes for estimating ψ_{kn}'s are summarized in Table 4.1. Here the MSEs of the MBB estimators of the four functionals ψ_{kn}, $k = 1, 2, 3, 4$, at various block lengths $\ell \in \{1, \ldots, 10\}$ are reported. The

TABLE 4.1. *Computation of the true optimal block sizes for the MBB estimation of the population parameters $\psi_{kn} : k = 1, 2, 3, 4$ of (4.2) under model (4.1) with $n = 80$, listed in parts (a)-(d) respectively. A double asterisk (**) in each part denotes the minimum value of the MSE.*

(a) ψ_{1n}

ℓ	Mean	Bias	SD	MSE
1	64.47265	-30.72935	10.29898	1050.3619
2	68.10531	-27.09669	13.79106	924.4239 **
3	69.83927	-25.36273	18.07921	970.1259
4	74.00986	-21.19214	21.92830	929.9571
5	76.38249	-18.81951	25.78332	1018.9536
6	76.96641	-18.23559	28.61257	1151.2162
7	77.37348	-17.82852	31.51760	1311.2153
8	77.12482	-18.07718	33.74574	1465.5595
9	77.20269	-17.99931	36.16197	1631.6632
10	76.15223	-19.04977	37.93431	1801.9055

(b) ψ_{2n}

ℓ	Mean	Bias	SD	MSE
1	0.5015600	-0.3050282	0.01291626	0.09320903 **
2	0.4996947	-0.3068935	0.01265911	0.09434389
3	0.5001387	-0.3064495	0.01379541	0.09410163
4	0.5007173	-0.3058709	0.01428579	0.09376107
5	0.4987853	-0.3078029	0.01483312	0.09496263
6	0.5003627	-0.3062255	0.01605235	0.09403176
7	0.5009587	-0.3056295	0.01579661	0.09365894
8	0.5007920	-0.3057962	0.01653342	0.09378467
9	0.5005200	-0.3060682	0.01699049	0.09396642
10	0.5009960	-0.3055922	0.01699534	0.09367543

(c) ψ_{3n}

ℓ	Mean	Bias	SD	MSE
1	-0.8471862	0.128348203	0.03843073	0.01795018
2	-0.8493528	0.126181608	0.04246156	0.01772478
3	-0.8574530	0.118081432	0.04813877	0.01626057
4	-0.8639622	0.111572175	0.04925145	0.01487406 **
5	-0.8699407	0.105593700	0.06120043	0.01489552
6	-0.8774479	0.098086509	0.07415318	0.01511966
7	-0.8866219	0.088912485	0.08711270	0.01549405
8	-0.8859254	0.089609010	0.09344112	0.01676102
9	-0.9081154	0.067418977	0.12055337	0.01907843
10	-0.9045705	0.070963860	0.12430460	0.02048750

(d) ψ_{4n}

ℓ	Mean	Bias	SD	MSE
1	0.5212083	-0.05931103	0.03524000	0.004759657
2	0.5274139	-0.05310544	0.03526084	0.004063515
3	0.5266830	-0.05383631	0.03987286	0.004488194
4	0.5284520	-0.05206733	0.04190782	0.004467272
5	0.5356206	-0.04489871	0.04444727	0.003991454 **
6	0.5353910	-0.04512829	0.04833994	0.004373312
7	0.5365192	-0.04400010	0.05429716	0.004884190
8	0.5368899	-0.04362945	0.05649139	0.005094806
9	0.5455250	-0.03499434	0.06642931	0.005637457
10	0.5400634	-0.04045590	0.06536473	0.005909228

MSE-values are based on 10,000 simulation runs, and for each data set, the bootstrap estimators were found by 500 replications. The minimum values of the MSE for each of the four functionals are marked with double asterisks (**). Thus, the 'true' optimal block sizes for estimating $\psi_{kn}, k = 1, 2, 3, 4$ are respecticely given by 2,1,4, and 5.

Table 4.2 gives the frequency distributions of the optimal block sizes for ψ_{kn}, $k = 1, 2, 3, 4$, selected by the Hall, Horowitz, and Jing (1995) method. The frequency -1 in Table 4.2 corresponds to the cases where the iterative process in the Hall, Horowitz, and Jing (1995) method failed to converge. Also, note that the value $\ell = 1$ was ruled out in the frequency distributions for the functionals ψ_{1n}-ψ_{4n}

TABLE 4.2. *Frequency distributions of the estimated optimal block sizes* $\bar{\ell}_n^0$ *for the functionals* ψ_{kn}, $k = 1, 2, 3, 4$, *produced by the Hall, Horowitz, and Jing (1995) method under model (4.1), listed in parts (a)-(d), respectively. Results are based on* 500 *simulation runs.*

(a) ψ_{1n}

ℓ	-1	2	3	4	6	7	9	10	12	13	15	18
Frequency	17	281	115	38	17	15	3	3	2	2	1	6

(b) ψ_{2n}

ℓ	-1	2	3	4	5	7	8
Frequency	90	295	62	18	25	8	2

(c) ψ_{3n}

ℓ	-1	2	3	4	5	7
Frequency	66	337	62	16	15	4

(d) ψ_{4n}

ℓ	-1	2	3	4	5	7	8	9
Frequency	119	258	57	24	34	5	2	1

by the extrapolation step (cf. (3.5)), which resulted in the high concentrations at the frquency 2 in all four cases. The frequency distributions of the optimal block sizes for ψ_{kn}, $k = 1, 2, 3, 4$, selected by the nonparametric plug-in method are given in Table 4.3. Although we report the frequencies for $\ell = 1$ in Table 4.3, as a rule of thumb, these should be merged with the next higher frequency $\ell = 2$, as resampling single observations (as opposed to resampling 'nontrivial' blocks of observations) at a time is not theoretically justified in the dependent case (cf. Singh (1981)).

From Tables 4.2 and 4.3, it follows that both the Hall, Horowitz, and Jing (1995) and the nonparametric plug-in methods have comparable performance. The tails of

TABLE 4.3. *Frequency distributions of the estimated optimal block sizes $\hat{\ell}_n^0$ for ψ_{kn}, $k = 1, 2, 3, 4$, produced by the nonparametric plug-in method under model (4.1), listed in parts (a)-(d), respectively. Results are based on* 500 *simulation runs.*

(a) ψ_{1n}

ℓ	1	2	3	4	5	6	7	8	9	10
Frequency	43	75	71	88	67	29	21	6	1	1

(b) ψ_{2n}

ℓ	1	2	3	4
Frequency	94	208	97	3

(c) ψ_{3n}

ℓ	1	2	3	4
Frequency	104	219	76	3

(d) ψ_{4n}

ℓ	1	2	3
Frequency	129	225	48

the frequency distributions are generally longer in the case of estimating the variance functional ψ_{1n} than in the other three cases. Furthermore, both methods are less effective in the case of the quantile functionals ψ_{3n} and ψ_{4n}. However, a more meaningful comparison of the performance of the two methods should be made in terms of the magnitudes of error incurred by the MBB estimators with the estimated block lengths compared to the MBB estimators based on true optimal block sizes. For more details on such comparisons and further discussions, see Lahiri, Furukawa, and Lee (2003) and Lahiri (2003).

5 Conclusions

In this article, we described two empirical methods for choosing the optimal block sizes for bootstrap estimation of various population parameters. The Hall, Horowitz, and Jing (1995) method uses the subsampling method to construct an estimate of the MSE function and minimizes it to produce the estimator of the optimal block length. In comparison, the nonparametric plug-in method uses the JAB method to generate the estimator of the optimal block length and has some attractive computational properties. An important feature of both methods is that they can be applied in various inference problems without requiring explicit expressions for the population parameters that appear in the theoretical optimal block lengths.

6 References

Bühlmann, P. (1994). Blockwise bootstrapped empirical process for stationary sequences. *Ann. Statist.* **22** 995-1012.

Bühlmann, P., and Künsch, H. R. (1999). Block length selection in the bootstrap for time series. *Computational Statistics and Data Analysis* **31** 295-310.

Carlstein, E. (1986). The use of subseries methods for estimating the variance of a general statistic from a stationary time series. *Ann. Statist.* **14** 1171-1179.

Carlstein, E., Do, K-A., Hall, P., Hesterberg, T., and Künsch, H. R. (1998). Matched-block bootstrap for dependent data *Bernoulli* **4** 305-328.

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

Efron, B. (1992). Jackknife-after-bootstrap standard errors and influence functions (with discussion). *J. Royal Stat. Soc. Ser. B* **54** 83-111.

Götze, F. and Hipp, C. (1983). Asymptotic expansions for sums of weakly dependent random vectors. *Zieb Wahr. verw Gebiete* **64** 211-239.

Götze, F. and Künsch, H.R. (1996). Second-order correctness of the blockwise bootstrap for stationary observations. *Ann. Statist.* **24** 1914 - 1933.

Hall, P. (1985). Resampling a coverage pattern. *Stoch. Proc. Appl.* **20** 231-246.

Hall, P. (1992). *The bootstrap and Edgeworth expansion.* Springer-Verlag Inc., New York.

Hall, P., Horowitz, J.L. and Jing, B-Y. (1995). On blocking rules for the bootstrap with dependent data. *Biometrika* **82** 561-574.

Künsch, H.R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.* **17** 1217-1261.

Lahiri, S.N. (1991). Second order optimality of stationary bootstrap. *Statist. Probab. Lett.* **11** 335-341.

Lahiri, S.N. (1996). On Edgeworth expansion and moving block bootstrap for studentized M-Estimators in multiple linear regression models. *J. Mult. Analysis* **56** 42- 59.

Lahiri, S. N. (1999). Theoretical comparisons of block bootstrap methods. *Ann. Statist.* **27** 386-404.

Lahiri, S. N. (2002). On the jackknife-after-bootstrap method for dependent data and its consistency properties. *Econometric Theory* **18** 79-98.

Lahiri, S. N. (2003). *Resampling methods for dependent data.* Springer-Verlag, New York, NY.

Lahiri, S.N., Furukawa, K., and Lee, Y-D. (2003). A nonparametric plug-in rule for selecting optimal block lengths for block bootstrap methods. *Preprint*, Department of Statistics, Iowa State University, Ames, IA.

Liu, R.Y. and Singh, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring the Limits of Bootstrap* (Eds.– R. Lepage and L. Billard). Wiley, New York, 225-248.

Naik-Nimbalkar, U.V. and Rajarshi, M.B. (1994). Validity of blockwise bootstrap for empirical processes with stationary observations. *Ann. Statist.* **22** 980-994.

Paparoditis, E. and Politis, D.N. (2001). Tapered block bootstrap. *Biometrika* **88** 1105-1119.

Politis, D. and Romano, J.P. (1992). A circular block resampling procedure for stationary data. In *Exploring the Limits of Bootstrap* (Eds.– R. Lepage and L. Billard), Wiley, New York, 263-270.

Politis, D. and Romano, J.P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *Ann. Statist.* **22** 2031-2050.

Politis, D. N., Romano, J. P., and Wolf, M. (1999). *Subsampling.* Springer-Verlag Inc., New York, NY.

Politis, D. and White, H. (2003). Automatic block length selection for the dependent bootstrap. *Preprint*, Department of Mathematics, University of California at San Diego, LaJolla, CA.

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