Matched–Block Bootstrap for Dependent Data

by

Edward Carlstein, Kim–Anh Do, Peter Hall,

Tim Hesterberg and Hans R. Künsch

Research Report No. 74

September 1996

Seminar für Statistik

Eidgenössische Technische Hochschule (ETH)

CH-8092 Zürich

Switzerland

MATCHED-BLOCK BOOTSTRAP FOR DEPENDENT DATA

Edward Carlstein[1] Kim-Anh Do[2,3] Peter Hall[3]
Tim Hesterberg[3,4] Hans R. Künsch[3,5]

SUMMARY. The block bootstrap for time series consists in randomly resampling blocks of consecutive values of the given data and aligning these blocks into a bootstrap sample. Here we suggest improving the performance of this method by aligning with higher likelihood those blocks which match at their ends. This is achieved by resampling the blocks according to a Markov chain whose transitions depend on the data. The matching algorithms we propose take some of the dependence structure of the data into account. They are based on a kernel estimate of the conditional lag one distribution or on a fitted autoregression of small order. Numerical and theoretical analyses in the case of estimating the variance of the sample mean show that matching reduces bias and, perhaps unexpectedly, has relatively little effect on variance. Our theory extends to the case of smooth functions of a vector mean.

KEY WORDS. Blocking methods, bootstrap, kernel methods, resampling, time series, variance estimation.

SHORT TITLE. Matched-block bootstrap.

AMS (1991) SUBJECT CLASSIFICATION. Primary 62G09, 62G15; Secondary 62E20.

---

[1] Department of Statistics, University of North Carolina, Chapel Hill, NC 27514, USA.
[2] Department of Social & Preventive Medicine, University of Queensland, Princess Alexandra Hospital, Woolloongabba, Queensland 4102, Australia.
[3] Centre for Mathematics and its Applications, Australian National University, Canberra, ACT 0200, Australia.
[4] Department of Mathematics, Franklin & Marshall College, Lancaster, PA 17604–3003, USA.
[5] Seminar für Statistik, ETH Zentrum, CH–8092 Zürich, Switzerland.

1 INTRODUCTION

In their classical form, as first proposed by Efron (1979), bootstrap methods are designed for application to samples of independent data. Under that assumption they implicitly produce an adaptive model for the marginal sampling distribution. During the last decade these approaches have been modified to suit the case of dependent data. Indeed, block bootstrap methods in that setting were introduced by Hall (1985), Carlstein (1986) and Künsch (1989). They involve implicitly computing empirical models for the general multivariate distribution of a stationary time series, or even a more general data sequence, under particularly mild assumptions on the process generating the data. The models are of course highly adaptive, or nonparametric, in the spirit of bootstrap methods. Since the introduction of the blockwise bootstrap, the method has been investigated in quite some detail. Shao and Yu (1993), Naik-Nimbalkar and Rajarshi (1994), Bühlmann (1994), Politis and Romano (1992) and Bühlmann and Künsch (1995) established consistency for a large number of statistics and processes generating the data. Recent work on the block bootstrap for distribution estimation includes contributions by Götze and Künsch (1993) and Lahiri (1993), showing that the block bootstrap can produce second-order correct estimators; and by Davison and Hall (1993), pointing out the need to carefully select the variance estimator when using a percentile-t version of the block bootstrap. Hall and Jing (1994), Hall, Horowitz and Jing (1995), and Bühlmann and Künsch (1994) have addressed the issue of block choice and related matters. Politis and Romano (1994, 1995) study modifications of the basic procedure.

The block bootstrap relies on producing a compromise between preserving the dependence structure of the original data and corrupting it by supposing that the data are independent. Blocks of data are resampled randomly with replacement from the original time series, and then a simulated version of the original process is assembled by joining the blocks together in random order. Although blocks of data are dependent in the original time series, they are independent in the bootstrap version. This causes bias in the bootstrap variance which can be large if the dependence in the data is strong. It is to be hoped that performance could be improved by matching the blocks in some way — that is, by using a block joining rule which in some sense favoured blocks that were *a priori* more likely to be close to one another. Our aim in the present paper is to analyse this procedure both numerically and theoretically. We show that in an important class of situations it does indeed produce improved performance.

There is a variety of ways in which matching can be effected. In Section 2 we present a number of matching rules which adapt to some extent to the nature of the data, for example by assuming a Markovian dependence or an autoregressive model. However, since the analysis of the matched blocks bootstrap is extremely difficult, we investigate mainly the case where blocks with similar values at the ends are paired. This is particularly appropriate when the data are generated by a continuous time process which is densely sampled so that the variance of the arithmetic mean decays at a slower rate than $O(n^{-1})$. Our results show that in this context, simple matching rules produce variance estimators that are less biased than, and have virtually the same variability as, those based on the ordinary, unmatched block bootstrap. The latter result is somewhat unexpected — one might have predicted that variance increases as a result of block matching, since it effectively introduces additional terms to the formula for the estimator. However, it turns out that the influence of those terms on variability is of second order.

Our numerical work and theory are in the context of estimating the variance of the sample mean. This case is of considerable practical as well as theoretical interest, not least because most statistics behave aymptotically as an arithmetic mean. There is no difficulty implementing the matched block bootstrap in a very wide range of settings, and indeed, theory in the case of a smooth function of the sample mean of a vector time series is very similar to that given here. However, the sample size necessary to reconcile asymptotic predictions with numerical results can be expected to grow as the smooth function becomes more complicated. In the case where the statistic is not a symmetric function of the data it is not clear whether one should compute the statistic directly on the bootstrap resample or use the "block of blocks" bootstrap idea of Künsch (1989), Politis and Romano (1992) and Bühlmann and Künsch (1995). Furthermore, we do not have a satisfactory theoretical account of the matched block bootstrap for distribution estimation, and so do not examine it in the present paper. However, it is likely that the matched-block bootstrap will also be efficacious in that setting.

Section 2 introduces a variety of matched-block bootstrap methods. Their asymptotic properties will be sketched in Section 3. These results are supported by simulation experiments in Section 4 and by rigorous arguments in Section 5. This leads to the main conclusion of this paper — that the matched-block bootstrap enhances performance by reducing the effect of bias, with relatively little influence on variance. In the case of a Markov process the bias reduction is by an order of magnitude, but in general it is by a constant factor.

We should mention here other methods which also reduce the bias. Künsch (1989) proposes a blockwise jackknife with smooth transition between observations left out and observations with full weight and similarly a weighted bootstrap. Politis and Romano (1995) suggest variance estimators that are essentially linear combinations of two block bootstrap estimators based on different block sizes.

2 METHODOLOGY FOR MATCHING BLOCKS

Given data $\mathcal{X} = \{X_i, 1 \le i \le n\}$ from a stationary time series, prepare blocks $\mathcal{B}_1, \cdots, \mathcal{B}_b$ each containing l (say) consecutive data values. There are two principal ways of doing this — using either non-overlapping blocks or overlapping blocks. Overlapping blocks could cause problems since there might be a strong tendency to match only neighbouring or nearly-neighbouring blocks. On the other hand from the case of non-matched blocks one would expect some gain in efficiency to result from using overlapping blocks (cf. Künsch, 1989; Hall, Horowitz and Jing 1995). In order to keep the presentation simple, we shall focus here on non-overlapping blocks, although our results have close parallels in the context of overlapping blocks. Hence we take b to be the integer part of n/l and $\mathcal{B}_i = \{X_{i1}, \cdots, X_{il}\}$ where $X_{ij} = X_{(i-1)l+j}$.

The matched blocks bootstrap constructs a Markov chain on the blocks with transition probabilities depending on the data $\mathcal{X}$. Specifically, if $\mathcal{B}_{i_1}, \cdots, \mathcal{B}_{i_j}$ are the first j blocks then the probability that the $(j+1)$'th block is $\mathcal{B}_{i_{j+1}}$ equals $p(i_j, i_{j+1})$. The first block is chosen according to the stationary distribution of the chain. As we shall see below, for our choices of the transition probabilities the stationary distirbution is close to the uniform. So we can start the chain also with the uniform distribution. The blocks obtained in this way are then put into a string $\mathcal{B}_{i_1}, \mathcal{B}_{i_2}, \cdots$. The first n values of this string constitute the bootstrap resample $\mathcal{X}^*$. If T is a function of n variables (representing the data) and $\hat{\theta} = T(\mathcal{X})$ is an estimator of an unknown parameter θ then, generally, $\hat{\theta}^* = T(\mathcal{X}^*)$ is its bootstrap version. The percentile form of the bootstrap estimates $\mathrm{var}\,(\hat{\theta})$ by $\mathrm{var}'(\hat{\theta}^*)$ and $P(\hat{\theta} - \theta \le t)$ by $P'(\hat{\theta}^* - \hat{\theta} \le t)$ where the prime denotes conditioning on the data $\mathcal{X}$. Centerings other than

$\hat{\theta}$ are possible. For example, if $\bar{X}$ denotes the sample mean then $E'(\bar{X}^*) \neq \bar{X}$ because the stationary distribution of the Markov chain will generally not be exactly uniform on the blocks, and not all observations appear in an equal number of blocks. However, the latter effect is only a boundary one, and the stationary distribution is in general quite close to being uniform.

Construction of the transition probabilities $p(i_1, i_2)$ is the essential part of our algorithm. Ideally we would do it in such a way that the bootstrap samples have properties similar to those of the original sample. On the other hand, there should be sufficient variability to produce a rich class of simulations, rather than simply reproducing the original sample. Our proposals achieve this by matching the blocks only through their values near the beginnings or ends of blocks. The simplest proposal is *kernel matching*, where

$$p(i_1, i_2) \propto \begin{cases} K\{(X_{i_1,l} - X_{i_2-1,l})/h\} & \text{if} \quad i_2 \neq 1 \\ K\{(X_{i_1+1,1} - X_{11})/h\} & \text{if} \quad i_2 = 1, i_1 l < n \\ 0 & \text{if} \quad i_2 = 1, i_1 l = n. \end{cases} \tag{2.1}$$

Here, K is a symmetric probability density and h is a bandwidth. The proportionality constant for each i_1 is determined by the requirement that for all i_1,

$$\sum_{i_2} p(i_1, i_2) = 1.$$

Note that this matching rule does not assume strong positive dependence for lag one since we match the last observation in $\mathcal{B}_{i_1}$ with the last observation in the block preceeding $\mathcal{B}_{i_2}$ in the original sample. Implicit in the matching rule is, however, an assumption that the dependence is mainly of Markovian character, since we use only the last observation in the block $\mathcal{B}_{i_1}$ to determine where $\mathcal{B}_{i_2}$ should start. In other words, the matching rule (2.1) corresponds to choosing the first element of $\mathcal{B}_{i_2}$, conditional on the last element of $\mathcal{B}_{i_1}$, according to the conditional distribution of X_i given X_{i-1}.

Alternatively, we can replace the observations X_i by their ranks. We call this *rank matching*. For later use we note that there are versions of rank matching where the stationary distribution is exactly uniform. For instance, if $n = bl$, we let R_i be the rank of X_{il} among $X_{1l}, \ldots, X_{bl}$ and $R_0 = R_b$. Now letting 1_A denote the indicator function of a set A, we set

$$p(j_1, j_2) = q(R_{j_1}, R_{j_2-1}),$$

where

$$q(i,j) = (2m+1)^{-1}(1_{\{|i-j|\leq m\}} + 1_{\{i+j\leq m+1\}} + 1_{\{i+j\geq 2b+1-m\}}).$$

It is easy to see that this defines a doubly stochastic matrix $(q(i,j))$, i.e. one where all row and column sums are equal to one.

Obviously we can also extend kernel and rank matching to the case where more than one observation (at the end of block i_1 and the block preceeding i_2) is used for the matching, in particular by taking products of kernels. This very quickly becomes impractical, however, because of the curse of dimensionality — either $p(i_1, i_2)$ is almost constant (if the bandwidth is large) or $p(i_1, i_1+1)$ dominates (if the bandwidth is small). An alternative procedure, *autoregressive matching*, takes into account $p < l$ observations at the ends of blocks. It is based on a fitted AR(p)–model, with coefficients $\hat{\phi}_1, \ldots, \hat{\phi}_p$ and distribution of the innovations given by $\hat{F}_\epsilon$. By iterating the defining equation of the model we produce matrices $A(\hat{\phi})$ and $B(\hat{\phi})$ such that

$$U_{i+p} = A(\hat{\phi})U_i + B(\hat{\phi})\,(\epsilon_{i+p}, \ldots, \epsilon_{i+1})',$$

where $U_i = (X_i, \ldots, X_{i-p+1})'$. This suggests the following algorithm. If the current block is $\mathcal{B}_{i_1}$, generate $\epsilon_1^*, \ldots, \epsilon_p^*$ by sampling independently from $\hat{F}_\epsilon$, and take the next block to be $\mathcal{B}_{i_2}$, where i_2 minimizes the L_1-norm of

$$(\epsilon_p^*, \ldots, \epsilon_1^*)' - B(\hat{\phi})^{-1}\,\{U_{(j-1)l+p} - A(\hat{\phi})\,U_{i_1 l}\}$$

over j. This amounts to choosing the first p values of the next block according to the fitted model, up to a discretization error. Autoregressive matching is thus a compromise between the AR-bootstrap (see Efron and Tibshirani, 1993) and the independent block bootstrap. Other ways to use a fitted AR-model for the matching are possible: we could for instance match that linear combination of values at the end of the blocks which predicts the average of future values best. This topic is left open for future research.

3 OVERVIEW OF LARGE-SAMPLE PROPERTIES

We consider using the block bootstrap to estimate the variance of the sample mean. For independent, nonoverlapping blocks we have

$$E\{\mathrm{var}'(\bar{X}^*)\} - \mathrm{var}\,(\bar{X}) \sim -\beta_1 = -2(nl)^{-1}\sum_{j=1}^{\infty} j\,\mathrm{cov}\,(X_0, X_j) \tag{3.1}$$

and

$$\operatorname{var}\{\operatorname{var}'(\bar{X}^*)\} \sim 2\,n^{-1}l\operatorname{var}(\bar{X})^2. \tag{3.2}$$

We argue that for a wide range of matching rules, (3.2) remains the same, but

$$E\{\operatorname{var}'(\bar{X}^*)\} - \operatorname{var}(\bar{X}) \sim -\beta_1 + \beta_2, \tag{3.3}$$

where β_2 is generally of the same sign as β_1 (for a non-repulsive process). In other words, block matching changes (and often reduces) the bias, but has relatively little effect on variance.

These properties will be derived rigorously in Section 5, for a slightly simplified procedure and a specific class of matching rules. We give here a simple recipe for calculating β_2 for general matching rules. Then we apply it to the rules introduced in Section 2.

The first step is to simplify the formula for the transition probabilities $p(i_1, i_2)$. We suppose that to a first approximation,

$$p(i_1, i_2) \sim b^{-1}\phi(U_{i_1}, V_{i_2-1}), \tag{3.4}$$

where U_i and V_i are functions of (X_{ij}) for j close to l. This reflects the fact that matching occurs mainly through the values near the ends of the blocks. The property $\sum_{i_2} p(i_1, i_2) = 1$ translates to

$$E\{\phi(u, V_i)\} \equiv 1. \tag{3.5}$$

For (3.2) to hold we need the stationary distribution of the chain to be approximately uniform. This means that

$$E\{\phi(U_i, v)\} \equiv 1. \tag{3.6}$$

Finally, the formula for β_2 is

$$\beta_2 = 2(nl)^{-1} \sum_{i=-\infty}^{0} \sum_{k=1}^{\infty} E\{E(X_i - \mu | U_0)\, E(X_k' - \mu | V_0')\, \phi(U_0, V_0')\}, \tag{3.7}$$

where $\mu = E(X_i)$ and $\{X_i'\}$ is an independent copy of $\{X_i\}$ (and V_i' is defined in terms of X_j'). In Section 5 it will become clear why this formula is to be expected.

Let us compute (3.4) and (3.7) for the matching rules of Section 2. For kernel matching we assume that X_i has density f. The law of large numbers suggests that the proportionality constant in (2.1) is

$$b \int K\{(X_{i_1 l} - y)/h\}\, f(y)\, dy.$$

Letting the bandwidth h tend to zero we obtain formally (3.7), with $U_i = V_i = X_{il}$ and $\phi(u,v) = f(u)^{-1}\delta(u-v)$, where δ denotes the Dirac delta function. Note that (3.5) and (3.6) are satisfied. Moreover,

$$\beta_2 = 2(nl)^{-1} \sum_{i=-\infty}^{0} \sum_{k=1}^{\infty} E\{E(X_i - \mu|X_0)\, E(X_k - \mu|X_0)\}. \tag{3.8}$$

For Gaussian processes (3.8) can be expressed with the covariance function. Moreover, if $\{X_i\}$ is a Markov process, then $\{X_i, i<0\}$ and $\{X_k, k>0\}$ are conditionally independent, given X_0. Thus,

$$E\{E(X_i - \mu|X_0)\, E(X_k - \mu|X_0)\} = E[E\{(X_i - \mu)\,(X_k - \mu)|X_0\}] = \mathrm{cov}\,(X_i, X_k),$$

whence

$$\beta_2 = 2(nl)^{-1} \sum_{i=-\infty}^{0} \sum_{k=1}^{\infty} \mathrm{cov}\,(X_i, X_k) = 2(nl)^{-1} \sum_{j=1}^{\infty} j\,\mathrm{cov}\,(X_0, X_j) = \beta_1.$$

This result and (3.3) show that for Markov processes, kernel matching reduces the bias of the bootstrap variance by an order of magnitude. That is understandable, since kernel matching relies on a Markovian assumption.

With rank matching the results are the same as those for kernel matching. So we turn to autoregressive matching. There we assume that the process $\{X_i\}$ is AR(p), that the innovations ϵ_i have a density g_1, and that the estimators $\hat{\phi}_j$ and $\hat{F}_\epsilon$ are consistent. Set $U_i = (X_{il}, \ldots, X_{i,l-p+1})'$, $V_i = (X_{i+1,p}, \ldots X_{i+1,1})'$ and

$$g_p(x_1, \ldots, x_p) = \prod_{i=1}^{p} g_1(x_i).$$

Denote the density of U_i by f. Then we expect (3.4) to hold with

$$\phi(u,v) = f(v)^{-1} g_p[B(\phi)^{-1}\{v - A(\phi)u\}].$$

Because $f(u)\, g_p[B(\phi)^{-1}\{v - A(\phi)u\}]$ is the joint density of (U_i, V_i) then (3.5) and (3.6) are satisfied, and (3.7) becomes

$$\beta_2 = 2(nl)^{-1} \sum_{i=-\infty}^{0} \sum_{k=1}^{\infty} E\{E(X_i - \mu|U_0)\, E(X_k - \mu|V_0)\}.$$

An AR(p) process is Markovian of order p, so we obtain by the same argument as before that again, $\beta_2 = \beta_1$. Therefore, autoregressive matching also reduces bias by an order of magnitude, provided the model behind the matching rule is correct.

4 NUMERICAL RESULTS

Our simulation study considered two models:

Model 1: First-order autoregressive processes, defined by $X_{t+1} = \rho X_t + (1-\rho^2)^{1/2}\,\epsilon_{t+1}$ for $\rho = 0.95$ and $\rho = 0.80$. The independent innovations ϵ_t were Normal $\mathcal{N}(0,1)$. We used the sample sizes $n = 200$, 1000 and 5000.

Model 2: Stationary Gaussian processes with covariance function $\gamma(t) = \text{cov}\,(X_s, X_{s+t}) = \exp(-c|t|^\alpha)$, where $c = 0.00015$ and $\alpha = 1.5$ or 1.95. The sample sizes were $n = 1000$ and 5000. We simulated these processes by using the algorithm developed by Wood and Chan (1994).

The results of the simulation study are summarized in Figures 1 and 2. The parameter of interest was $\sigma^2_{\bar{X}} = \text{var}\,(\bar{X})$. The mean squared errors depicted in the figures were obtained by averaging over 500 independent simulations, using $B = 500$ bootstrap resamples within each simulation. The bandwidth chosen for each simulation was $h = b^{-1/5}\hat{\sigma}$, where $\hat{\sigma}$ was the sample standard deviation of the simulated sequence. (Except for the fact that the constant is 1 rather than 1.06, this is the "equivalent Normal density" prescription for bandwidth selection; see Silverman (1986, p. 45).)

As before, we write l for block length. Let MSE_m and MSE_o denote the mean squared errors of the matched and ordinary block bootstrap methods, respectively. The following features emerged from our simulation study. Most importantly, the minimum mean squared error was consistently smaller under the matched block bootstrap than under the ordinary block bootstrap. As predicted by our theory, the block length at which the minimum occurred tended to be smaller for the matched block bootstrap. For example, in the case of Model 1 with $(\rho, n) = (0.80, 1000)$, the smallest value of MSE_m was 4.0×10^{-6} and occurred at $l = 5$, whereas the smallest value of MSE_o was 7.1×10^{-6} and occurred at $l = 35$. (Both values of l are correct to the nearest multiple of 5.) The bias terms were consistently negative for both methods. Note particularly that these trends were also apparent in Model 1 where the dependence was relatively long-range. In the case of the ordinary block bootstrap, optimal block length showed a marked tendency to increase with increasing range of dependence. This trend was not reflected in the matched block method, however.

Similar results, not reported here, were obtained in simulations from a third model, a moving average with negative coefficient: $X_t = a\epsilon_{t-1} + b\epsilon_t$, where $a = -b = 2^{-1/2}$. This has $\gamma(1) < 0$ and $\gamma(i) = 0$ for $i \geq 2$, and is "repulsive" in the sense that no covariance other than the variance is positive and thus $\beta_1 < 0$. But (3.8) shows that $\beta_2 = \beta_1/2$. So kernel matching is expected to reduce bias which is confirmed by our simulations. Indeed, the only essential difference noted between results in this and the other two models was that here, all biases were positive. As before, the minimum mean squared error was smaller for the matched block bootstrap than for the ordinary block bootstrap.

5 THEORETICAL RESULTS

We now turn to rigorous derivation of results (3.2) and (3.3). The technical details of theory for block-matching are particularly arduous. To keep them in manageable, succinct form we treat a somewhat abstract version of the procedure that we earlier discussed in Sections 2 – 4. For simplicity we assume that $n = bl$ for integers b (the number of blocks) and l (the length of each block). We consider the case where time series data $\{X_i\}$ are derived by sampling a continuous process with a sampling frequency which may increase with n. So, let $\{Y(t),\ t \in (0,\infty)\}$ denote a stationary stochastic process in continuous time, implying that $\mu \equiv E\{Y(t)\}$ does not depend on t, and $\gamma(t) \equiv \text{cov}\,\{Y(s), Y(s+t)\}$ does not depend on s. Let $\lambda = \lambda(n)$ represent a sequence of positive constants possibly diverging to infinity as n increases, and put $X_i = Y(i/\lambda)$. The strength of dependence of the process $\{X_i\}$ increases with increasing λ. Indeed, the variance of the sample mean is of order $O(\lambda/n)$ (see below), and so long-range dependence might be considered to be characterized by the case where λ increases with sample size.

Our assumptions on the process Y are as follows:

(C_{1A}) Y is t_0-dependent for some $t_0 > 0$, meaning that the sigma-fields $\mathcal{F}(0,s)$ and $\mathcal{F}(s+t_0,\infty)$ generated by $\{Y(u),\ u \in (0,s)\}$ and $\{Y(u),\ u \in (s+t_0,\infty)\}$, respectively, are independent for each $s > 0$;

(C_{1B})

$$E|Y(t)|^\alpha < \infty \quad \text{for some } \alpha > 8, \quad \text{to be determined later.}$$

Condition (C_{1A}) simplifies our arguments, but modified versions of our results hold in the case where Y is mixing with geometrically decreasing mixing rate.

Next we set down our assumptions on the block matching algorithm. Remember that $p(j_1, j_2)$ for $1 \le j_1, j_2 \le b$ is the data-dependent probability that the next block is $\mathcal{B}_{j_2}$, given that the current block is $\mathcal{B}_{j_1}$. Put $V_i = (X_{i1}, \ldots, X_{ir})$ and $U_i = (X_{i,l-r+1}, \ldots, X_{il})$, the first r and the last r values in block i respectively. We impose the following conditions:

($\mathrm{C_{2A}}$) for all j_1, j_2,

$$p(j_1, j_2) = \psi(U_{j_1}, V_{j_2}; V_j, j \ne j_2),$$

where ψ is nonnegative and symmetric in the last $b-1$ arguments, and $r = O(\lambda)$;

($\mathrm{C_{2B}}$) for all j_1,

$$\sum_{j_2=1}^{b} p(j_1, j_2) = 1;$$

($\mathrm{C_{2C}}$) for some $\epsilon > 0$,

$$\sup_j E\{p(j, j+1)\} = O(b^{-\epsilon});$$

($\mathrm{C_{2D}}$) for any j_1, j_2, j_3, j_4 with $j_3 \ne j_2$, $j_4 \ne j_2$, there exists $p'(j_1, j_2, j_3, j_4) \in [0, 1]$ which depends only on U_{j_1} and $V_j, j \ne j_3, j_4$ such that for some $\epsilon > 0$ and all $1 \le q < \infty$,

$$\sup_{j_1, j_2, j_3, j_4} \|p(j_1, j_2) - p'(j_1, j_2, j_3, j_4)\|_q = O(b^{-1-\epsilon}),$$

where $||.||_q$ denotes the L_q norm;

($\mathrm{C_{2E}}$) for some $\epsilon > 0$,

$$\operatorname{ess\,sup}_{j_2 \ne j_1+1} \; E\{p(j_1, j_2) | \, V_{j_1}; \mathcal{B}_j, j \ne j_1, j_1+1\} = O(b^{-\epsilon}).$$

For example, we might define

$$\begin{aligned} p_1(j_1, j_2) &= \prod_{k=0}^{r-1} K\{(X_{j_1, l-k} - X_{j_2, r-k})/h\}, \qquad (5.1) \\ p(j_1, j_2) &= p_1(j_1, j_2) \Big/ \left\{ \sum_{j=1}^{b} p_1(j_1, j) \right\}, \end{aligned}$$

where $K \ge 0$ denotes a bounded and compactly supported kernel function and h is a bandwidth satisfying $h = O(b^{\epsilon})$ and $b^{1-\epsilon} h \to \infty$ for some $\epsilon > 0$; and $1 \le r = r(n) = o(\lambda)$. In the event that the denominator in the definition of $p(j_1, j_2)$ vanishes, define $p(j_1, j_2)$ to equal b^{-1} for each value of j_2. Conditions ($\mathrm{C_2}$) may be verified in this setting, for a wide variety of processes including polynomial functions of Gaussian processes whose covariance γ satisfies ($\mathrm{C_{1A}}$). In this setting the approximating probability $p'(j_1, j_2, j_3, j_4)$ in condition

(C_{2D}) may be constructed by removing from the denominator in the definition of $p(j_1, j_2)$ a finite number of terms $p_1(j_1, j)$, so as to achieve the desired independence. Furthermore, condition (C_{2E}) is an immediate consequence of the compact support of K and of the conditions imposed on h. Note that by way of contrast to rule (2.1), rule (5.1) now assumes strong positive dependence for neighbouring values, which is natural in the context of dense sampling of a continuous process.

Of course, many alternative prescriptions of p are possible, still satisfying conditions (C_2). In particular, there is considerable latitude for varying the block representatives that are compared via the kernel function in the definition of p_1 at (5.1).

Let $\bar{X}$ and $\bar{X}^*$ denote sample means of the data $\mathcal{X}$ and resampled data $\mathcal{X}^*$, respectively; and let

$$\sigma^2 = \sigma^2(n) = \operatorname{var}(\bar{X}) = n^{-1}\left\{\gamma(0) + 2\sum_{j=1}^{n-1}(1 - n^{-1}j)\,\gamma(j/\lambda)\right\}$$

represent the variance of the sample mean. The matched-block bootstrap estimator of σ^2 is given by $\hat{\sigma}^2 = \operatorname{var}'(\bar{X}^*)$, where the prime denotes conditioning on $\mathcal{X}$. To appreciate the size of the quantity that we are estimating, note that if $\lambda \to L$ as $n \to \infty$, where $0 < L \leq \infty$, then

$$\sigma^2 \sim \begin{cases} n^{-1}\left\{\gamma(0) + 2\sum_{j=0}^{\infty}\gamma(j/L)\right\} & \text{if } L < \infty \\ 2(\lambda/n)\int_0^\infty \gamma(t)\,dt & \text{if } L = \infty\,. \end{cases}$$

Therefore, σ^2 is of size λ/n.

Let the stationary distribution on the block indices $(1, \cdots, b)$ be $\pi_1, \cdots, \pi_b$. Assume that the blocks $\mathcal{B}_{i_j}$ are produced with the chain in this stationary state, and put $\bar{X}' = \sum \pi_i \bar{X}_i$, $\bar{X}_i = l^{-1}\sum_j X_{ij}$. Then $E'(\bar{X}^*) = \bar{X}'$ and $\hat{\sigma}^2 = b^{-2}(S_1 + 2S_2)$, where

$$\begin{aligned} S_1 &= b\sum_{i=1}^{b}\pi_i(\bar{X}_i - \bar{X}')^2\,, \\ S_2 &= \sum_{j=1}^{b-1}\sum_{i=1}^{b-j} E'\{(\bar{X}_j^* - \bar{X}')(\bar{X}_{i+j}^* - \bar{X}')\} \\ &= \sum_{i=1}^{b-1}(b-i)\,E'\{(\bar{X}_1^* - \bar{X}')(\bar{X}_{i+1}^* - \bar{X}')\} \\ &= \sum_{i=1}^{b-1}(b-i)\sum_{j_1=1}^{b}\sum_{j_2=1}^{b}\pi_{j_1}p(i; j_1, j_2)\,(\bar{X}_{j_1} - \bar{X}')(\bar{X}_{j_2} - \bar{X}') \end{aligned}$$

and $p(i; j_1, j_2)$ denotes the i-step transition probability in the Markov chain of blocks $(p(1; j_1, j_2) = p(j_1, j_2))$.

If the stationary distribution of the block-matching rule is approximately uniform and is reached after two steps, then to a good approximation, $S_j \approx T_j$ where

$$T_1 = \sum_{i=1}^{b} (\bar{X}_i - \bar{X})^2 \,, \qquad T_2 = \sum_{j_1=1}^{b} \sum_{j_2=1}^{b} p(j_1, j_2)\,(\bar{X}_{j_1} - \bar{X})(\bar{X}_{j_2} - \bar{X})\,.$$

This suggests an alternative variance estimator,

$$\tilde{\sigma}^2 = b^{-2}(T_1 + 2T_2)\,.$$

We shall describe theory for this quantity. We believe that $\tilde{\sigma}^2$ contains the essential features of $\hat{\sigma}^2$, for the following reasons. We showed earlier that the stationary distribution is uniform for a version of rank matching, and that it is approximately uniform in other cases since (3.6) is satisfied. That the stationary distribution is reached after two steps is plausible because V_{j_2} and U_{j_2} are independent if l is large. Hence, the two terms $p(j_1, j_2)$ and $p(j_2, j_3)$ are essentially independent.

As the theorem below shows, the leading term in an expansion of bias is of size $(nl)^{-1}\lambda^2$, and equals $-\beta_1 + \beta_2 + o(\lambda^2/nl)$ where

$$\begin{aligned}
\beta_1 &\equiv 2(nl)^{-1} \sum_{i=1}^{\infty} i\gamma(i/\lambda) = \left(\lambda^2/nl\right) c_1 + o\left(\lambda^2/nl\right)\,, \\
c_1 &\equiv 2\int_0^\infty t\gamma(t)\,dt\,, \\
\beta_2 &\equiv 2E\{p(1,3)(\bar{X}_1 - \mu)(\bar{X}_3 - \mu)\}\,.
\end{aligned}$$

We shall also show that β_2 is typically of the same order as β_1.

Theorem 1. *Assume conditions (C_1) on the process Y, and (C_2) on the matching rule, with $\alpha > \max(8, 4/\epsilon)$, $\lambda = o(l)$ and $l = o(n)$. Then*

$$E(\tilde{\sigma}^2) = \sigma^2 - \beta_1 + \beta_2 + O\left(\lambda n^{-2} l\right) + o\left\{\lambda^2/(nl)\right\}\,, \tag{5.2}$$

$$\mathrm{var}\,(\tilde{\sigma}^2) = 2n^{-1} l \sigma^4 + o\left\{\lambda^2 n^{-3} l + \lambda^4 (nl)^{-2}\right\}\,. \tag{5.3}$$

Since either β_1^2 or $n^{-1}l\sigma^4$ dominates each remainder term then it is always true that

$$E\left\{(\tilde{\sigma}^2 - \sigma^2)^2\right\} \sim (\beta_1 - \beta_2)^2 + 2n^{-1}l\sigma^4.$$

If in addition $l = o\{(n\lambda)^{1/2}\}$ and $n\lambda^2 = O(l^3)$ then

$$E(\tilde{\sigma}^2) - \sigma^2 \sim -\beta_1 + \beta_2, \quad \mathrm{var}\,(\tilde{\sigma}^2) \sim 2n^{-1} l \sigma^4. \tag{5.4}$$

The last result represents an analogue of (3.2) and (3.3).

In order to compute the exact order of β_2 and its leading term, we need stronger assumptions on the matching rules. A class of different rules is covered by

Theorem 2. *Assume in addition to the conditions of the previous theorem that for some* $\epsilon_1, \epsilon_2 > 0$, *and all* $1 < q < \infty$,

$$\|b\,p(1,3) - I(|X_{1l} - X_{31}| \le b^{-\epsilon_1})\,\{P(|X_{1l} - X_{31}| \le b^{-\epsilon_1}|X_{1l})\}^{-1}\|_q = O(b^{-\epsilon_2})\,.$$

Suppose too that $\lambda \to \infty$, Y *is an almost surely continuous process and that* $Y(t)$ *has a continuous density with respect to Lebesgue measure. Then*

$$\beta_2 = \gamma_2\lambda^2/(nl) + o\{\lambda^2/(nl)\}, \tag{5.5}$$

where

$$\gamma_2 = \sum_{i=-l+1}^{0}\sum_{k=0}^{l-1} E\{E[\{Y(i/\lambda) - \mu\}|Y(0)]\,E[\{Y(k/\lambda) - \mu\}|Y(0)]\}\lambda^{-2}.$$

If in addition the process Y *is Gaussian, with* $\gamma(t) - \gamma(0) = O(|t|^\epsilon)$ *for some* $\epsilon > 0$, *then*

$$\gamma_2 \sim c_2 = 2\gamma(0)^{-1}\left\{\int_0^\infty \gamma(t)\,dt\right\}^2. \tag{5.6}$$

Remark 1. Observe that the first-order contributions to squared bias and variance are of sizes $\lambda^4(nl)^{-2}$ and $\lambda^2 n^{-3} l$, respectively. Therefore the optimal block length is of size $(n\lambda^2)^{1/3}$. Result (5.4) holds for such values of l.

Remark 2. As in Section 3, $\gamma_2 \sim c_1$ if Y is a Markov process.

Remark 3. For a general stationary distribution π_i,

$$\mathrm{var}\,(S_1) \sim 2\sum \pi_i^2\sigma^4,$$

so for (5.3) it seems necessary to have $\sum \pi_i^2 \sim b^{-1}$. This implies, via the Cauchy-Schwarz inequality, that the stationary distribution is approximately uniform.

Remark 4. Without condition ($\mathrm{C_{1A}}$) we should replace the indices 1 and 3 in β_2 by two indices j_1, j_2 with $|j_1 - j_2|$ tending to infinity. Since then $\mathcal{B}_{j_1}$ and $\mathcal{B}_{j_2}$ become independent, we obtain (3.7) from (3.4).

The effect of block matching on the bias may be studied most easily for Gaussian processes. Figure 3 depicts the asymptotic value of the ratio of the biases for matched and non-matched blocks, $1 - c_2/c_1$, for the covariance functions $\gamma(t) = \exp(-c|t|^\alpha)$ (In this example we do not adhere to the technical assumption (C_{1A})). Here $0 < \alpha \leq 2$, and the larger α the smoother the process is. This example shows that the reduction in bias can be substantial.

To appreciate that block-matching in terms of nearness of block ends is counter-productive for a time series with a considerable amount of repulsion, note that because c_2 is always positive, block matching by nearness of block ends exacerbates the bias problem when $c_1 < 0$. To be specific, consider the case

$$\gamma(t) = \begin{cases} (1 - |t|)\cos(\omega t) & \text{if } |t| \leq 1 \\ 0 & \text{otherwise}, \end{cases}$$

where ω is a parameter of the process. The value of c_1 for this covariance function is $4\omega^{-3}\sin\omega - 2\omega^{-2}(1+\cos\omega)$, which is negative for many choices of ω. For $\omega \approx \pi$ the bias of the matched block estimator is substantially larger than the bias of the non-matched block bootstrap, see Figure 4. Similar behaviour is observed with other covariance functions that have negative parts.

APPENDIX: PROOFS OF THEOREMS.

Step 1: Preliminaries. Assume that $E(Y) = 0$, and define

$$T_3 = \sum_{j=1}^{b} \bar{X}_j^2, \quad T_4 = \sum_{j_1=1}^{b}\sum_{j_2=1}^{b} p(j_1, j_2)\,\bar{X}_{j_1}\,\bar{X}_{j_2}, \quad T_5 = \sum_{j_1=1}^{b}\sum_{j_2=1}^{b} p(j_1, j_2)\,\bar{X}_{j_2}.$$

Since $\sum_{j_2} p(j_1, j_2) = 1$ then $T_2 = T_4 - \bar{X}T_5$. Therefore,

$$\tilde{\sigma}^2 = b^{-2}(T_3 + 2T_4) - b^{-1}\,\bar{X}^2 - 2b^{-2}\,\bar{X}T_5\,.$$

It is straightforward to prove that

$$E\left(\bar{X}^4\right) = O\left(b^{-2}\,l^{-2}\,\lambda^2\right),$$

and we shall show in Step 2 that for some $\zeta > 0$,

$$E\left(\bar{X}^2\,T_5^2\right) = O\left\{b^{1-\zeta}\,(\lambda/l)^2 + b^{2-\zeta}\,(\lambda/l)^4\right\}. \tag{A.1}$$

With similar but simpler arguments one can show that

$$E\left(\bar{X}\,T_5\right) = O\left\{\lambda l^{-1} + b^{1-\zeta}\,(\lambda/l)^2\right\}.$$

In Step 3 we show that

$$E\,(T_4) = b^2 E\,\{p(1,3)\bar{X}_1\bar{X}_3\} + O(\lambda/l) + O(b^{1-\zeta}(\lambda/l)^2). \tag{A.2}$$

Using an argument similar to that in Step 2 it may be proved that

$$\mathrm{var}\,(T_4) = o\left\{b^4(\lambda/n)^2\,\left(b^{-1} + l^{-2}\,\lambda^2\right)\right\}.$$

From the bootstrap with independent blocks we know that

$$E\,(T_3) = b\sigma^2(l) = b^2\sigma^2(n) - b^2\beta_1 + O\left\{(\lambda/l)^2\right\},$$

$$\mathrm{var}\;(T_3) \sim 2b^3\sigma^4(n) = O\left\{b(\lambda/l)^2\right\}.$$

These results, together with the Cauchy-Schwarz inequality and the fact that $b = n/l$, imply (5.2) and (5.3).

Step 2: Proof of (A.1). Observe that

$$E\left(\bar{X}^2\,T_5^2\right) = b^{-2}\sum\nolimits^{(6)} E\{p(j_1,j_2)\,p(j_3,j_4)\bar{X}_{j_2}\,\bar{X}_{j_4}\,\bar{X}_{j_5}\,\bar{X}_{j_6}\} = b^{-2}\,l^{-4}S\,, \tag{A.3}$$

where

$$S \equiv \sum\nolimits^{(10)} E\{p(j_1,j_2)\,p(j_3,j_4)X_{j_2k_1}X_{j_4k_2}X_{j_5k_3}X_{j_6k_4}\}\,,$$

the six-fold sum $\sum^{(6)}$ is over vectors $(j_1,\cdots,j_6) \in \{1,\cdots,b\}^6$, and the ten-fold sum $\sum^{(10)}$ is over those vectors and also over $(k_1,\cdots,k_4) \in \{1,\cdots,l\}^4$.

We bound S by considering a number of different configurations of the vectors $(j_1,\cdots,j_6)$ and $(k_1,\cdots,k_4)$. We call k_i a *boundary index* if $k_i \le r + t_0\,\lambda$ or $k_i \ge l - r - t_0\,\lambda + 1$, and an interior index otherwise. The cases identified below cover all distinct configurations up to isomorphisms. Since there is only a bounded number of the latter then we do not treat them here.

Case I: $k_1,\cdots,k_4$ are all interior indices. Here the term

$$E\{p(j_1,j_2)\,p(j_3,j_4)\,X_{j_2k_1}X_{j_4k_2}X_{j_5k_3}X_{j_6k_4}\} \tag{A.4}$$

factorizes into the product of $E\{p(j_1,j_2)\,p(j_3,j_4)\}$ and $E(X_{j_2k_1}X_{j_4k_2}X_{j_5k_3}X_{j_6k_4})$. The second factor equals zero unless (a) $j_2 = j_4 = j_5 = j_6$, or (b) $j_2 = j_5 \ne j_4 = j_6$, or (c) $j_2 = j_4 \ne j_5 = j_6$, or one of the bounded number of possibilities isomorphic to these obtains. In subcase (a) the sum over $k_1,\cdots,k_4$ contributes a term of order $l^2\lambda^2$, and the sum over j_1, j_2 and j_3 contributes another $O(b^2)$. (Note that the sum of $p(j_1, j_2)$ over its second index is identically 1.) Since the sums are in multiple then these two contributions should be multiplied together, and so the contribution to S obtained by summing the term at (A.4) over indices corresponding to subcase (a) is $b^2 l^2 \lambda^2$. The argument in subcase (b) is similar, with identical orders of magnitude arising from summation over $k_1, \cdots, k_4$ and over $j_1, \cdots, j_4$. Therefore, the contribution to S that arises in subcase (b) is again $O(b^2 l^2 \lambda^2)$.

The contribution to S from subcase (c) is

$$\sum_{j_1} \sum_{j_2} \sum_{j_3} \sum_{j_5} E\{p(j_1, j_2)\, p(j_3, j_2)\}\, O\left(l^2 \lambda^2\right) . \tag{A.5}$$

In bounding the expectation we may suppose that $j_2 \neq j_3 + 1$, since the contrary case may be treated more simply. (There, the number of sums in (A.5) is effectively only three, not four.) Under this assumption we may define $p'(j_1, j_2, j_3 + 1, j_3 + 1)$ as in condition $(C_{2\mathrm{D}})$. Then,

$$\begin{aligned} E\{p(j_1, j_2)\, p(j_3, j_2)\} \quad &\leq \quad E\{|p(j_1, j_2) - p'(j_1, j_2, j_3 + 1, j_3 + 1)|\, p(j_3, j_2)\} \\ &\qquad + E\{p'(j_1, j_2, j_3 + 1, j_3 + 1)\, p(j_3, j_2)\}. \end{aligned}$$

By the symmetry in condition ($\mathrm{C}_{2\mathrm{A}}$), $E\{p(j_3, j_2)\} \leq 1/(b-1)$. Using ($\mathrm{C}_{2\mathrm{D}}$) and choosing $q > 2/\epsilon$, the first term on the right is seen to be bounded by

$$\|p(j_1, j_2) - p'(j_1, j_2, j_3, j_3)\|_q\, E\{p(j_3, j_2)\}^{1-(1/q)} = O\left(b^{-1-\epsilon}\, b^{-1+(1/q)}\right) = O\left(b^{-1-(\epsilon/2)}\right) .$$

Moreover, by ($\mathrm{C}_{2\mathrm{D}}$) and ($\mathrm{C}_{2\mathrm{E}}$), the second term on the right is bounded by

$$\begin{aligned} & E[p'(j_1, j_2, j_3 + 1, j_3 + 1) E\{p(j_3, j_2) | V_{j_3}; \mathcal{B}_j,\, j \neq j_3, j_3 + 1\}] \\ & \quad \leq \quad C b^{-\epsilon}\, E\{p'(j_1, j_2, j_3 + 1, j_3 + 1)\} \\ & \quad \leq \quad C b^{-\epsilon}\, E\{|p(j_1, j_2) - p'(j_1, j_2, j_3 + 1, j_3 + 1)| + p(j_1, j_2)\} \leq C' b^{-1-\epsilon} . \end{aligned}$$

Therefore, the expectation in (A.5) equals $O(b^{-1-(\epsilon/2)})$, and so the quantity at (A.5) equals $O(b^{3-(\epsilon/2)} l^2 \lambda^2)$.

Combining the results from subcases (a) to (c) we see that the contribution to S that arises from case I equals $O(b^{3-\zeta}\, l^2\, \lambda^2)$, for some $\zeta > 0$.

Case II: $k_1, \cdots, k_4$ are all boundary indices. Defining

$$\pi = \pi(j_1, \cdots, j_4) = p(j_1, j_2)\, p(j_3, j_4) ,$$

the term at (A.4) becomes $E(\pi X_{j_2k_1} X_{j_4k_2} X_{j_5k_3} X_{j_6k_4})$. We consider separately the subcases (a) j_5 or j_6 belongs to $\{j_1-1, j_1, j_2, j_3-1, j_3, j_4\}$, and (b) all other situations. In subcase (a) we bound the term by

$$\left\{E\left(\pi^{\alpha/(\alpha-4)}\right)\right\}^{(\alpha-4)/\alpha} \left(E|X_{j_2k_1} X_{j_4k_2} X_{j_5k_3} X_{j_6k_4}|^{\alpha/4}\right)^{4/\alpha} .$$

In subcase (a) the number of values (j_5, j_6) is $O(b)$ uniformly in $(j_1, \cdots, j_4)$, so summing over j_5 and j_6 gives a contribution $O(b)$. By Hölder's inequality

$$\begin{aligned} \sum_{j_1 \cdots j_4} \|\pi\|_{\alpha/(\alpha-4)} &\leq \left\{\sum_{j_1 \cdots j_4} E\left(\pi^{\alpha/(\alpha-4)}\right)\right\}^{(\alpha-4)/\alpha} \left(\sum_{j_1 \cdots j_4} 1\right)^{4/\alpha} \\ &\leq \left\{\sum_{j_1 \cdots j_4} E(\pi)\right\}^{(\alpha-4)/\alpha} b^{16/\alpha} = b^{2+8/\alpha}. \end{aligned}$$

Combining these results we see that the total contribution to S in subcase (a) equals $O(b^{3+(8/\alpha)}\lambda^4)$. Taking $\alpha > 8$ thus ensures that this contribution does not exceed

$$O\left(b^{4-\zeta}\lambda^4\right) \tag{A.6}$$

for some $\zeta > 0$.

Next we treat subcase (b). Let k_3 and k_4 be distant $O(\lambda)$ from $(j_5-1)l+1$ and $(j_6-1)l+1$, respectively. Define

$$\pi' = \pi'(j_1, \cdots, j_6) = p'(j_1, j_2, j_5, j_6)\, p'(j_3, j_4, j_5, j_6) .$$

Then we have, in view of the independence of $X_{j_5k_3} X_{j_6k_4}$ and $\pi' X_{j_2k_1} X_{j_4k_2}$,

$$\begin{aligned} &|E(\pi X_{j_2k_1} X_{j_4k_2} X_{j_5k_3} X_{j_6k_4}) - E(\pi X_{j_2k_1} X_{j_4k_2})\, E(X_{j_5k_3} X_{j_6k_4})| \\ &\quad \leq |E\{(\pi - \pi') X_{j_2k_1} X_{j_4k_2} X_{j_5k_3} X_{j_6k_4}\}| \\ &\qquad + E\left(X_1^2\right) |E\{(\pi - \pi') X_{j_2k_1} X_{j_4k_2}\}| . \end{aligned} \tag{A.7}$$

By Hölder's inequality, the first term on the right-hand side of (A.7) is bounded by $\|\pi - \pi'\|_{\alpha/(\alpha-4)}\, \|X_1\|_\alpha^4$. Since

$$\begin{aligned} \pi - \pi' \;=\;& p(j_1, j_2)\, \{p(j_3, j_4) - p'(j_3, j_4, j_5, j_6)\} \\ & + p(j_3, j_4)\, \{p(j_1, j_2) - p'(j_1, j_2, j_5, j_6)\} \\ & - \{p(j_1, j_2) - p'(j_1, j_2, j_5, j_6)\}\, \{p(j_3, j_4) - p'(j_3, j_4, j_5, j_6)\} , \end{aligned}$$

then the triangle inequality, Hölder's inequality and condition (C_{2D}) imply that for any $\xi > 1$,

$$\begin{aligned}\|\pi - \pi'\|_{\alpha/(\alpha-4)} &= O(b^{-1-\epsilon}) \left\{\|p(j_1, j_2)\|_{\xi\alpha/(\alpha-4)} + \|p(j_3, j_4)\|_{\xi\alpha/(\alpha-4)}\right\} \\ &\quad + O\left(b^{-2(1+\epsilon)}\right).\end{aligned}$$

Similarly, the second term on the right-hand side of (A.7) is bounded by

$$\begin{aligned}\|\pi - \pi'\|_{\alpha/(\alpha-2)} \|X_1\|_\alpha^2 \|X_1\|_2^2 &= O\left(b^{-1-\epsilon}\right) \left\{\|p(j_1, j_2)\|_{\xi\alpha/(\alpha-2)}\right. \\ &\quad \left. + \|p(j_3, j_4)\|_{\xi\alpha/(\alpha-2)}\right\} + O\left(b^{-2(1+\epsilon)}\right).\end{aligned}$$

If k_3 is within distance $O(\lambda)$ of $j_5 l$ then we argue as above but with the definition of π' altered to $p'(j_1, j_2, j_5 + 1, j_6)\, p'(j_3, j_4, j_5 + 1, j_6)$. Thus, the bounds just derived hold for any of the terms arising in subcase (b) of case II. Moreover, $E(X_{j_5 k_3} X_{j_6 k_4}) = 0$ unless $|j_5 - j_6| \le 1$. This, together with (A.7) and the bounds above, produce the following bound for the contribution to S from case II, subcase (b):

$$\begin{aligned}O(b\lambda^4) \sum_{j_1,\cdots,j_4} \|\pi(j_1, j_2, j_3, j_4)\|_{\alpha/(\alpha-2)} &+ O\left(b^{4-(1+\epsilon)}\,\lambda^4\right) \sum_{j_1, j_2} \|p(j_1, j_2)\|_{\xi\alpha/(\alpha-4)} \\ &+ O\left(b^{6-2(1+\epsilon)}\,\lambda^4\right).\end{aligned}$$

Arguing as in the derivation of the bound at (A.6) we see that this equals

$$O\left(b^{3+(4/\alpha)}\,\lambda^4 + b^{5+(4\xi/\alpha)-\epsilon-(1/\xi)}\,\lambda^4 + b^{4-2\epsilon}\,\lambda^4\right),$$

which, since we made the assumption $\alpha > 4/\epsilon$, may be rendered of the order at (A.6) by choosing $\xi > 1$ sufficiently close to 1.

Adding the bounds from subcases (a) and (b) we see that the total contribution to S from terms considered under case II is of the order at (A.6).

Case III: Three k_i*'s are boundary indices and the other is interior.* Here the contribution to S is identically zero.

The methods used to derive the bounds in Cases IV–VII below are somewhat different, and are given only in barest outline here. Although the bounds are identical, none of the cases is isomorphic to another.

Case IV: k_1, k_2 *are boundary indices and* k_3, k_4 *are interior.* The contribution is identically zero unless $j_5 = j_6$, and there the contributions from the sums over

$$(k_3, k_4),\ (k_1, k_2),\ j_5 \text{ and } (j_1, j_2, j_3, j_4)$$

are respectively $O(l\lambda)$, $O(\lambda^2)$, $O(b)$ and $O(b^{2(\alpha-2)/\alpha}\, b^{4(2/\alpha)})$. Multiplying them together we see that the total contribution to S is $O(b^{3+(4/\alpha)}\, l\lambda^3)$. Because $\alpha > 8$ and the geometric mean is bounded by the arithmetic mean, we have that $b^{3+(4/\alpha)} l\lambda^3 = O(b^{3-\zeta}\, l^2\lambda^2 + b^{4-\zeta}\, \lambda^4)$ for some $\zeta > 0$.

Case V: k_3, k_4 are boundary indices and k_1, k_2 are interior. The contribution to S is identically zero unless $j_2 = j_4$, and the contribution from the latter source is $O(b^{4-\zeta}\, l\lambda^3)$ for some $\zeta > 0$, using an argument similar to that employed to treat subcase (c) of Case I.

Case VI: k_1, k_3 are boundary indices and k_2, k_4 are interior indices. The contribution to S is identically zero unless $j_4 = j_6$, and the contribution from the latter source is $O(b^{3+(4/\alpha)}\, l\lambda^3)$.

Case VII: k_4 is a boundary index and the others are all interior. The contribution to S is identically zero unless $j_2 = j_4 = j_5$, and the contribution from the latter source is $O(b^{3+(1/\alpha)}\, l\lambda^3)$.

Case VIII: k_2 is a boundary index and the others are all interior. The contribution to S is identically zero unless $j_2 = j_5 = j_6$, and the contribution from the latter source is $O(b^{2+(2/\alpha)}\, l\lambda^3)$.

Now we add the bounds derived in each of the eight cases. Because $\alpha > 8$ and the geometric mean is bounded by the arithmetic mean, $b^{3+(4/\alpha)} l\lambda^3 = O(b^{3-\zeta}\, l^2\lambda^2 + b^{4-\zeta}\, \lambda^4)$ for some $\zeta > 0$. Hence we obtain that for some $\zeta > 0$,

$$S = O\left(b^{3-\zeta}\, l^2\lambda^2 + b^{4-\zeta}\, \lambda^4\right) .$$

Result (A.1) now follows from (A.3).

Step 3: Calculation of $E(T_4)$. Note that by symmetry, $E\{p(j_1, j_2)\, \bar{X}_{j_1} \bar{X}_{j_2}\}$ is the same for any $j_1 < b$, $j_2 < b$, $|j_2 - j_1| > 1$. By an argument similar to that in Step 2 we may show that for any j_1, j_2,

$$|E\{p(j_1, j_2)\, \bar{X}_{j_1}\, \bar{X}_{j_2}\}| = O(\lambda/l)\, E\{p(j_1, j_1)\}\, \delta_{j_1, j_2} + O\left\{(\lambda/l)^2\right\}\, \|p(j_1, j_2)\|_{\alpha/(\alpha-2)}.$$

Since $p(j_1, j_2) \le 1$ then

$$\|p(j_1, j_2)\|_{\alpha/(\alpha-2)} \le [E\{p(j_1, j_2)\}]^{(\alpha-2)/\alpha} .$$

Using the fact that $E\{p(j_1, j_2)\} \le 1/(b-1)$ if $j_2 \ne j_1 + 1$, and condition (C_{2C}) if $j_2 = j_1 + 1$, we obtain for $\alpha > 8$,

$$E(T_4) - b^2 E\{p(1,3)\, \bar{X}_1\, \bar{X}_3\} = O(\lambda/l) + O\left\{(\lambda/l)^2 b^{1-3\epsilon/4}\right\},$$

which is (A.2).

Step 4: Proof of (5.5) and (5.6). Let Y, Y_1, Y_2 be independent processes with identical laws, and put

$$Z_j = \sum_{i=0}^{l} Y_j(i/\lambda)\,.$$

It is easy to see that under the conditions of Theorem 2,

$$\beta_2 = b^{-1}\,l^{-2}\,E[E\{Z_1 Z_2 | Y_1(0) = Y_2(0)\}] + o\left(\lambda^2/bl^2\right),$$

which is (5.5). Finally, for a Gaussian process, $E\{Y(i/\lambda)|Y(0)\} = \gamma(i/\lambda)/\gamma(0)\,Y(0)$, which gives (5.6).

REFERENCES

BÜHLMANN, P. (1994). Blockwise bootstrapped empirical processes for stationary sequences. *Ann. Statist.* **22**, 995–1012.

BÜHLMANN, P. AND KÜNSCH, H.R. (1994). Block length selection in the bootstrap for time series. Research Report No. 72, Seminar für Statistik, ETH Zurich.

BÜHLMANN, P. AND KÜNSCH, H.R. (1995). The blockwise bootstrap for general parameters of a stationary time series. *Scand. J. Statist.*, **22**, 35 – 54.

CARLSTEIN, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *Ann. Statist.* **14**, 1171–1179.

DAVISON, A.C. AND HALL, P. (1993). On Studentizing and blocking methods for implementing the bootstrap with dependent data. *Aust. J. Statist.* **35** 215–224.

EFRON, B. (1979). Bootstrap methods: another look at the jackknife. (With discussion.) *Ann. Statist.* **7**, 1–26.

EFRON, B. AND TIBSHIRANI, R.J. (1993). An Introduction to the Bootstrap. Chapman and Hall, New York.

GÖTZE, F. AND KÜNSCH, H.R. (1993). Second order correctness of the blockwise bootstrap for stationary observations. *Ann. Statist.*, under revision.

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

HALL, P. AND JING, B. (1994). On sample re-use methods for dependent data. Research Report No. SR8–94, Centre for Mathematics and its Applications, Australian National University.

HALL, P., HOROWITZ, J.L. AND JING, B. (1995). On blocking rules for the block bootstrap with dependent data. *Biometrika* **82**, 561–574.

KÜNSCH, H.R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.* **17**, 1217–1241.

LAHIRI, S.N. (1993). On Edgeworth expansions and the moving block bootstrap for Studentized M-estimators in multiple linear regression models. *J. Multivar. Anal.*, to appear.

NAIK-NIMBALKAR, U.V. AND RAJARSHI, M.B. (1994). Validity of blockwise bootstrap for empirical processes with stationary observations. *Ann. Statist.* **22**, 980–994.

POLITIS, D.N. AND ROMANO, J.P. (1992). A general resampling scheme for triangular arrays of α-mixing random variables with application to the problem of spectral density estimation. *Ann. Statist.* **20**, 1985–2007.

POLITIS, D.N. AND ROMANO, J.P. (1994). The stationary bootstrap. *J. Amer. Statist. Assoc.* **89**, 1303–13.

POLITIS, D.N. AND ROMANO, J.P. (1995). Bias-corrected nonparametric spectral estimation. *J. Time Series Anal.* **16**, 67–103.

SHAO, Q.M. AND YU, H. (1993). Bootstrapping the sample means for stationary mixing sequences. *Stoch. Proc. Appl.* **48**, 175–190.

SILVERMAN, B.W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London.

WOOD, A.T.A. AND CHAN, G. (1994). Simulation of stationary Gaussian processes. *J. Comp. Graph. Statist.* **31**, 409–432.

Figure 1: MSE comparisons for Model 1. Column panels correspond to $\rho = 0.8$, 0.95 respectively, and row panels to $n = 200$, 1000, 5000 respectively. The solid curve represents the matched block bootstrap, the dotted curve the ordinary block bootstrap.

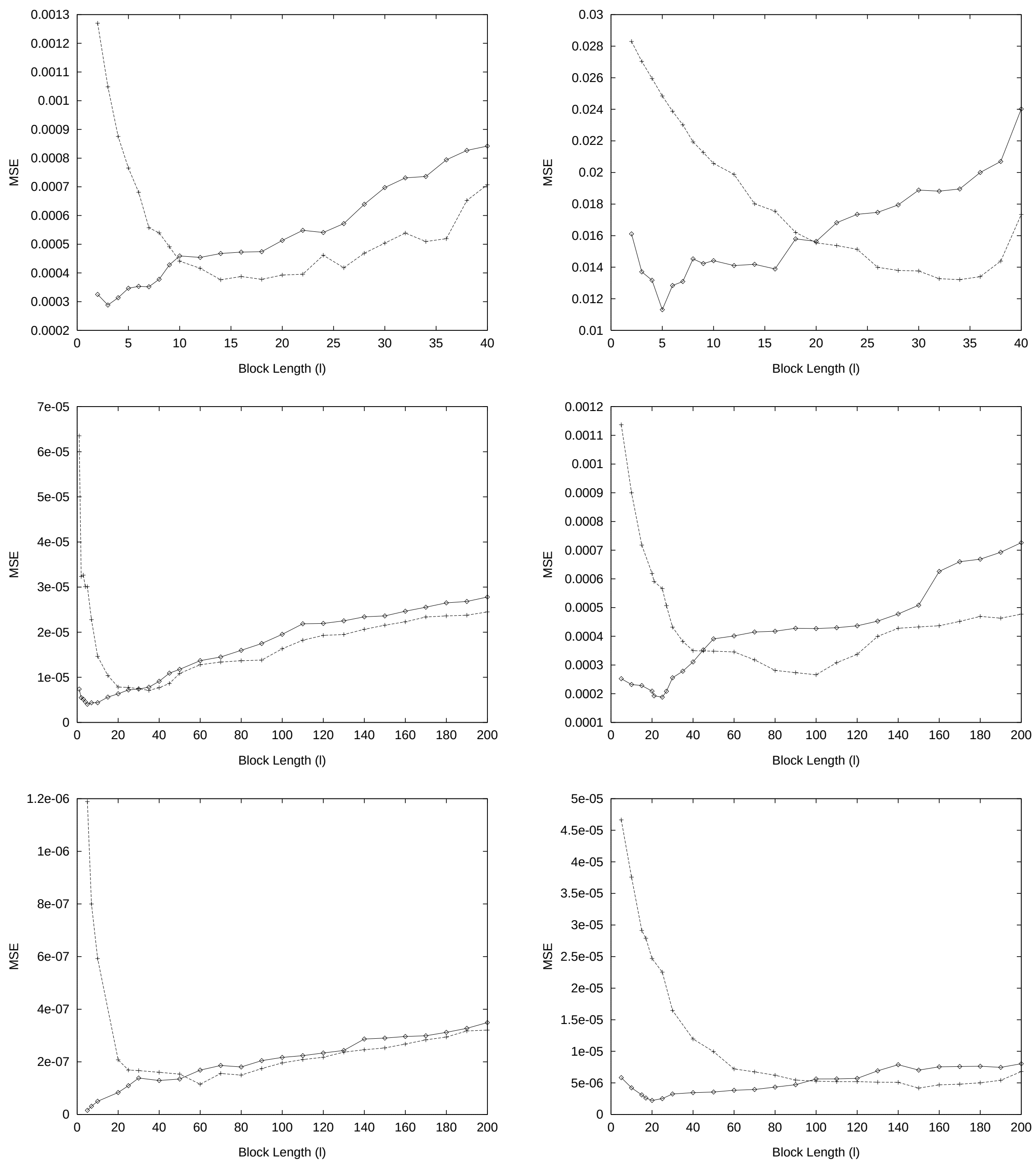

Figure 2: MSE comparisons for Model 2 when $c = 0.00015$. Column panels correspond to $\alpha = 1.5$, 1.95 respectively; row panels to $n = 1000$, 5000 respectively. The solid curve represents the matched block bootstrap, the dotted curve the ordinary block bootstrap.

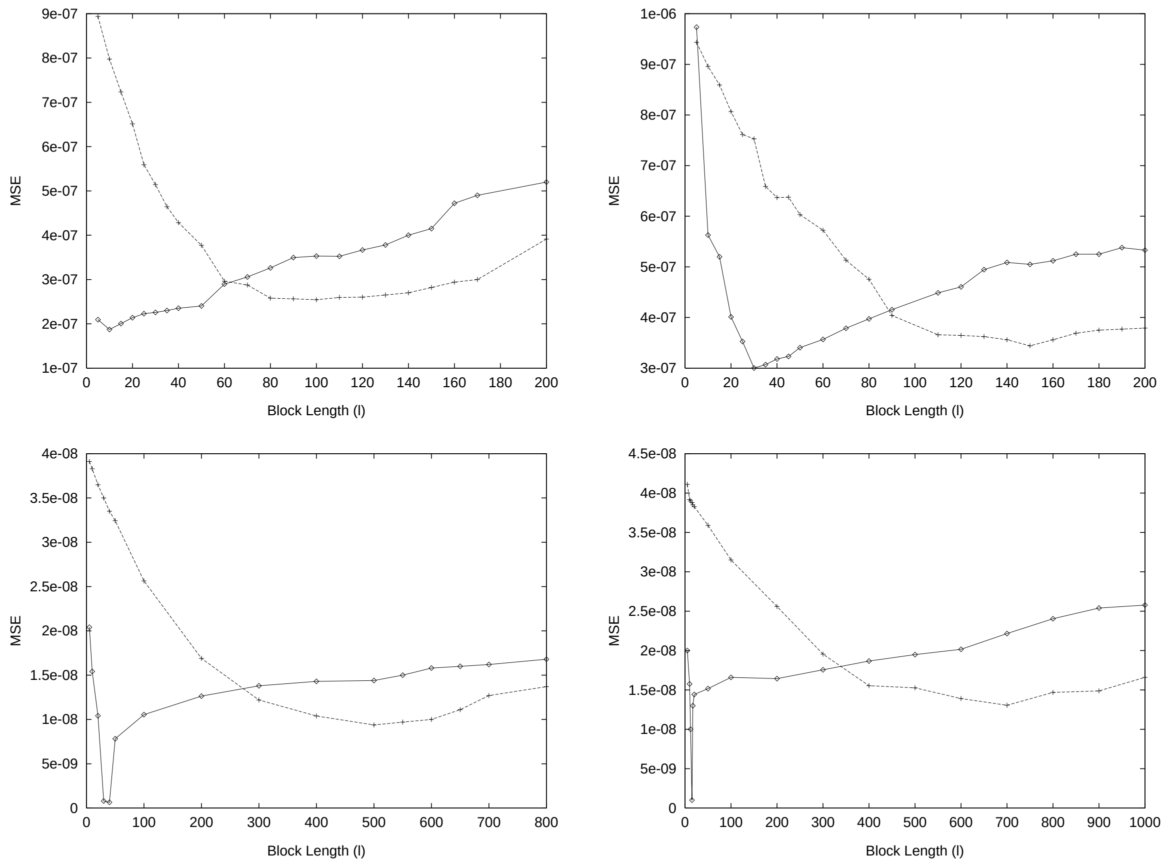

Figure 3: The figure depicts values of the ratio $1 - c_2/c_1$, the ratio of the bias of the matched to the bias of the non-matched block bootstrap, for Gaussian processes with $\gamma(t) = \exp(-c|t|^{\alpha})$, where $0 < \alpha \leq 2$; the ratio is independent of c.

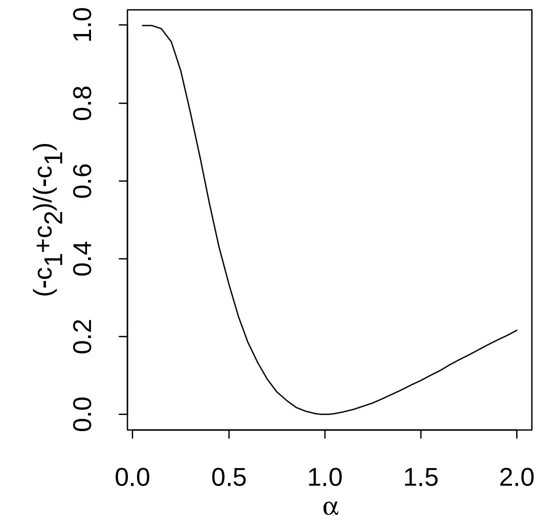

Figure 4: The figure depicts values of $-c_1$ and of $(-c_1 + c_2)$, the leading terms of the bias of the non-matched and matched block bootstrap, respectively, for the case $\gamma(t) = (1 - |t|)\cos(\omega t)1_{\{|t|<1\}}$. This covariance kernel exhibits repulsion if ω is sufficiently large, so that the non-matched block bootstrap can have positive bias.