Statistical Science
2003, Vol. 18, No. 2, 219–230

The Impact of Bootstrap Methods on Time Series Analysis

Dimitris N. Politis

Abstract. Sparked by Efron's seminal paper, the decade of the 1980s was a period of active research on bootstrap methods for independent data—mainly i.i.d. or regression set-ups. By contrast, in the 1990s much research was directed towards resampling dependent data, for example, time series and random fields. Consequently, the availability of valid nonparametric inference procedures based on resampling and/or subsampling has freed practitioners from the necessity of resorting to simplifying assumptions such as normality or linearity that may be misleading.

Key words and phrases: Block bootstrap, confidence intervals, linear models, resampling, large sample inference, nonparametric estimation, subsampling.

1. INTRODUCTION: THE SAMPLE MEAN OF A TIME SERIES

Let $X_1, \ldots, X_n$ be an observed stretch from a strictly stationary time series $\{X_t, t \in \mathbf{Z}\}$; the assumption of stationarity implies that the joint probability law of $(X_t, X_{t+1}, \ldots, X_{t+k})$ does not depend on t for any $k \geq 0$. Assume also that the time series is weakly dependent; that is, the collection of random variables $\{X_t, t \leq 0\}$ is approximately independent of $\{X_t, t \geq k\}$ when k is large enough. An example of a weak dependence structure is given by m-dependence under which $\{X_t, t \leq 0\}$ is (exactly) independent of $\{X_t, t \geq k\}$ whenever $k > m$; independence is just the special case of 0-dependence.

Due to the dependence between the observations, even the most basic methods involved in applied statistical work suddenly become challenging; an elementary such example has to do with estimating the unknown mean $\mu = EX_t$ of the time series. The sample mean $\bar{X}_n = n^{-1} \sum_{t=1}^{n} X_t$ is the obvious estimator; however—and here immediately the difficulties crop up—it is not the most efficient. To see why, consider the regression $X_t = \mu + \varepsilon_t$, which also serves as a definition for the ε_t process; it is apparent that $\bar{X}_n$ is the ordinary least squares estimator of μ in this model. However, because of the dependence in the errors ε_t, the best linear unbiased estimator (BLUE) of μ is instead obtained by a generalized least squares argument; consequently, $\hat{\mu}_{\mathrm{BLUE}} = (\mathbf{1}'\Gamma_n^{-1}\mathbf{1})^{-1}\mathbf{1}'\Gamma_n^{-1}\mathbf{X}$, where $\mathbf{X} = (X_1, \ldots, X_n)'$, $\mathbf{1} = (1, \ldots, 1)'$ and Γ_n is the (unknown) covariance matrix of the vector $\mathbf{X}$ with i, j element given by $\gamma(i-j) = \mathrm{Cov}(X_i, X_j)$.

It is immediate that $\hat{\mu}_{\mathrm{BLUE}}$ is a weighted average of the X_t data, (i.e., $\hat{\mu}_{\mathrm{BLUE}} = \sum_{i=1}^{n} w_i X_i$); the weights w_i are a function of the (unknown) covariances $\gamma(s)$, $s = 0, 1, \ldots, n-1$. For example, under the simple AR(p), that is, autoregressive of order p, model,

$$(X_t - \mu) = \phi_1(X_{t-1} - \mu) + \cdots + \phi_p(X_{t-p} - \mu) + Z_t, \tag{1}$$

where $Z_t \sim$ i.i.d. $(0, \tau^2)$, it is easy to see that $w_j = w_{n-j+1}$ for $j = 1, \ldots, p$, $w_{p+1} = w_{p+2} = \cdots = w_{n-p}$ and $\sum_{i=1}^{n} w_i = 1$; calculating $w_1, \ldots, w_p$ in terms of $\gamma(\cdot)$ is feasible but cumbersome—see Fuller (1996). For instance, in the special case of $p = 1$, we can calculate that $w_1 = w_2/(1 - \rho(1))$, where $\rho(s) = \gamma(s)/\gamma(0)$ is the autocorrelation sequence.

Nevertheless, it is apparent that, at least in the AR(p) model above, the difference of the w_i weights from the constant weights in $\bar{X}_n$ is due mostly to the end effects, that is, the first and last p weights.

Dimitris N. Politis is Professor of Mathematics and Adjunct Professor of Economics, University of California, San Diego, La Jolla, California 92093-0112 (e-mail: dpolitis@ucsd.edu).

Thus, it may be reasonable to conjecture that, for a large sample size n, those end effects will have a negligible contribution. This is, in fact, true; as shown in Grenander and Rosenblatt (1957) for a large class of weakly dependent processes [not limited to the AR(p) models], the sample mean $\bar{X}_n$ is asymptotically efficient. Therefore, estimating μ by our familiar, easy-to-use, sample mean $\bar{X}_n$ is fortuitously justified.

The problem, however, is further complicated if one wants to construct an estimate of the standard error of $\bar{X}_n$. To see this, let $\sigma_n^2 = \mathrm{Var}(\sqrt{n}\bar{X}_n)$ and note that $n\sigma_n^2$ is just the sum of all the elements of the matrix Γ_n, that is, $\sigma_n^2 = \sum_{s=-n}^{n}(1 - |s|/n)\gamma(s)$. Under the usual regularity conditions,

$$\sigma_\infty^2 := \lim_{n\to\infty} \sigma_n^2 = \sum_{s=-\infty}^{\infty} \gamma(s) = 2\pi f(0),$$

where $f(w) = (2\pi)^{-1}\sum_{s=-\infty}^{\infty} e^{iws}\gamma(s)$, for $w \in [-\pi, \pi]$, is the spectral density function of the $\{X_t\}$ process. It is apparent now that the simple task of standard error estimation is highly nontrivial under dependence, as it amounts to nonparametric estimation of the spectral density at the origin.

One may think that estimating γ and plugging into the above formula would give a consistent estimator of the infinite sum $\sum_{s=-\infty}^{\infty}\gamma(s)$. Estimators of $\gamma(s)$ for $|s| < n$ can certainly be constructed; $\tilde{\gamma}(s) = (n - s)^{-1}\sum_{t=1}^{n-s}(X_t - \bar{X}_n)(X_{t+s} - \bar{X}_n)$ and $\hat{\gamma}(s) = n^{-1}\sum_{t=1}^{n-s}(X_t - \bar{X}_n)(X_{t+s} - \bar{X}_n)$ are the usual candidates. Note that $\tilde{\gamma}(s)$ is unbiased but becomes quite unreliable for large s because of increasing variability; at the extreme case of $s = n - 1$, there is no averaging at all in $\tilde{\gamma}(s)$. Thus, $\hat{\gamma}(s)$ is preferable for two reasons: (a) it trades in some bias for a huge variance reduction by using the assumption of weak dependence [which implies that $\gamma(s) \to 0$ as $s \to \infty$] to shrink the $\tilde{\gamma}(s)$ toward 0 for large s; and (b) $\hat{\gamma}(s)$ is a nonnegative definite sequence, that is, it guarantees that $\sum_{|s|<n}\hat{\gamma}(s) \geq 0$. To understand the last claim, note that $\sum_{|s|<n}\hat{\gamma}(s) = 2\pi T(0)$, where $T(w)$ is the periodogram defined by

$$T(w) = (2\pi n)^{-1}\left|\sum_{s=1}^{n} e^{iws}(X_s - \bar{X}_n)\right|^2,$$

which, however, is well known to be inconsistent as an estimator of $f(w)$; although the bias of $T(w)$ is negligible [of order $O(1/n)$], its variance is approximately constant and does not approach 0—see, for example, Brockwell and Davis (1991). Consequently, the plug-in idea $\sum_{|s|<n}\hat{\gamma}(s)$ is inconsistent as well; intuitively, it is seen that $\hat{\gamma}(s)$ is still unreliable for large s and more shrinking toward 0 is necessary.

The above observations suggest two different ways of achieving consistent spectral estimation:

A. Split the data series $X_1, \ldots, X_n$ into (overlapping) blocks of size b, that is, $(X_1, \ldots, X_b)$, $(X_2, \ldots, X_{b+1})$ and so on. Calculate the periodogram $T_i(w)$ from block $(X_i, \ldots, X_{i+b-1})$, for $i = 1, \ldots, q$ (where $q = n - b + 1$), and then let $\bar{T}(w) = q^{-1}\sum_{i=1}^{q} T_i(w)$ be the average of those short periodograms. It follows that the bias of $\bar{T}(w)$ is equal to the bias of each of the $T_i(w)$, that is, it is of order $O(1/b)$, while its variance is $O(b/n)$ due to the averaging. [To see that $\mathrm{Var}(\bar{T}(w)) = O(b/n)$, note that periodograms calculated from nonoverlapping blocks are approximately independent; there are approximately n/b such nonoverlapping blocks within the data stretch $X_1, \ldots, X_n$. By a Cauchy–Schwarz argument, it is immediate that the variance of $\bar{T}(w)$ cannot be bigger than the variance of the average of just those n/b approximately independent, nonoverlapping periodograms which is $O(b/n)$.] Thus, consistency of $\bar{T}(w)$ ensues provided

(2) $\qquad b \to \infty \quad$ as $n \to \infty$ but with $b/n \to 0$.

This blocking idea was actually one of the very first spectral estimation schemes proposed in a pioneering paper by Bartlett (1946).

B. Since more shrinking of $\hat{\gamma}(s)$ toward 0 is advisable for the purposes of convergence of their sum, it comes as no surprise that a popular method of estimating $f(w)$ is given by means of truncated/weighted sums:

$$\hat{f}_{b,\lambda}(w) = (2\pi)^{-1}\sum_{s=-b}^{b} \lambda(s/b)e^{iws}\hat{\gamma}(s), \tag{3}$$

where the lag window λ is some (well-chosen) function on $[-1, 1]$. Consistency is achieved again under (2); in particular, the condition $b = o(n)$ effectively keeps the unreliable $\hat{\gamma}(s)$—the ones for large s—from entering in and spoiling the convergence of the sum in (3).

Interestingly, the above two methods of spectral estimation turn out to be equivalent to each other—as well as to the third popular method that entails smoothing the periodogram; see Brockwell and Davis (1991). For example, Bartlett's $\bar{T}(w)$ is approximately equivalent to $\hat{f}_{b,\lambda}(w)$ using the triangular lag window

$\lambda_B(x) = 1 - |x|$. Although the constant lag window $\lambda_C(x) = 1$ also leads to consistent estimation under (2), improved accuracy is achieved via the family of flat-top lag windows introduced in Politis and Romano (1995); the simplest representative of the flat-top family is the trapezoidal lag window $\lambda_T(x) = \min\{1, 2(1 - |x|)\}$, which can be thought of as a compromise between λ_C and λ_B.

2. BLOCK RESAMPLING AND SUBSAMPLING

As is well known, the bootstrap scheme pioneered in the seminal paper of Efron (1979) was geared toward independent data. As a matter of fact, it was quickly recognized by Singh (1981) that if one applies the i.i.d. bootstrap to data that are dependent, inconsistency follows; see also the early paper by Babu and Singh (1983). Because of the total data "scrambling" induced by the i.i.d. bootstrap, all dependence information is lost. Consequently, the i.i.d. bootstrap estimate of σ_n^2 typically does not converge to the correct limit given by $\sigma_\infty^2 = \sum_{s=-\infty}^{\infty} \gamma(s)$; rather, it converges to $\gamma(0)$, that is, the i.i.d. bootstrap fails to give valid standard error estimates in the dependent case.

2.1 Subsampling

Subsampling is probably one of the most intuitive methods of valid statistical inference. To see why subsampling works in the time series case, consider the aforementioned consistent estimator of σ_∞^2 due to Bartlett, that is, $2\pi \bar{T}(0)$ or, equivalently, $2\pi \hat{f}_{b,\lambda_B}(0)$. Recall that $2\pi \hat{f}_{b,\lambda_B}(0) = \sum_{s=-b}^{b}(1 - |s|/b)\hat{\gamma}(s)$, which is actually a simple, plug-in estimator of $\sigma_b^2 = \mathrm{Var}(\sqrt{b}\sum_{i=1}^{b} X_i) = \sum_{s=-b}^{b}(1 - |s|/b)\gamma(s)$. Intuitively, $2\pi \hat{f}_{b,\lambda_B}(0)$ is estimating σ_b^2 and not σ_n^2; consistency of $2\pi \hat{f}_{b,\lambda_B}(0)$ is achieved only because both σ_b^2 and σ_n^2 tend to σ_∞^2 under (2)—and therefore also $\sigma_b^2 - \sigma_n^2 \to 0$.

In other words, it seems that we cannot directly estimate σ_n^2 or σ_∞^2, and we have to content ourselves with estimating σ_b^2. But an immediate, empirical estimator of $\sigma_b^2 = \mathrm{Var}(\sqrt{b}\sum_{i=1}^{b} X_i)$ can be constructed from the many size-b sample means that can be extracted from our data $X_1, \ldots, X_n$. To formalize this notion, let $\bar{X}_{i,b} = b^{-1}\sum_{t=i}^{i+b-1} X_t$ be the sample mean of the ith block and let $\hat{\sigma}_{b,\mathrm{SUB}}^2$ denote the sample variance of the (normalized) subsample values $\sqrt{b}\bar{X}_{i,b}$ for $i = 1, \ldots, q$. The simple estimator $\hat{\sigma}_{b,\mathrm{SUB}}^2$ is the subsampling estimator of variance; it is consistent for σ_∞^2 under (2) which is not surprising in view of that fact that it, too, is equivalent to the ubiquitous Bartlett estimator!

Nevertheless, the beauty of the subsampling methodology is its extreme generality: let $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$ be an arbitrary statistic that is consistent for a general parameter θ at rate a_n, that is, for large n, $a_n(\hat{\theta}_n - \theta)$ tends to some well-defined asymptotic distribution J; the rate a_n does not have to equal $\sqrt{n}$, and the distribution J does not have to be normal—we do not even need to know its shape, just that it exists. Let $\hat{\theta}_{i,b} = \hat{\theta}_b(X_i, \ldots, X_{i+b-1})$ be the subsample value of the statistic computed from the ith block. The subsampling estimator of J is $\hat{J}_{b,\mathrm{SUB}}$ defined as the empirical distribution of the normalized (and centered) subsample values $a_b(\hat{\theta}_{i,b} - \hat{\theta}_n)$ for $i = 1, \ldots, q$.

The consistency of $\hat{J}_{b,\mathrm{SUB}}$ for general statistics under minimal conditions—basically involving a weak dependence condition and (2)—was shown in Politis and Romano (1992c, 1994b); consequently, confidence intervals for θ can immediately be formed using the quantiles of $\hat{J}_{b,\mathrm{SUB}}$ instead of the quantiles of the (unknown) J. A closely related early work is the paper by Sherman and Carlstein (1996) where the subsampling distribution of the (unnormalized) subsample values is put forth as a helpful diagnostic tool.

If a variance estimator for $a_n\hat{\theta}_n$ is sought, it can be constructed by the sample variance of the normalized subsample values $a_b\hat{\theta}_{i,b}$ for $i = 1, \ldots, q$; consistency of the subsampling estimator of variance for general statistics was shown by Carlstein (1986) under some uniform integrability conditions. It should be pointed out that if there is no dependence present, then the ordering of the data is immaterial; the data can be permuted with no information loss. Thus, in the i.i.d. case, there are $\binom{n}{b}$ subsamples of size b from which subsample statistics can be recomputed; in this case, subsampling is equivalent to the well-known delete-d jackknife (with $d = n - b$) of Wu (1986) and Shao and Wu (1989); see Politis, Romano and Wolf (1999) for more details and an extensive list of references.

Subsampling in the i.i.d. case is also intimately related to the i.i.d. bootstrap with smaller resample size b, that is, where our statistic is recomputed from $X_1^*, \ldots, X_b^*$ drawn i.i.d. from the empirical distribution of the data $X_1, \ldots, X_n$; the only difference is that subsampling draws b values *without* replacement from the dataset $X_1, \ldots, X_n$ while the bootstrap draws *with* replacement.

The possibility of drawing an arbitrary resample size was noted as early on as in Bickel and Freedman (1981); nevertheless, the research community focused for some time on the case of resample size n since that choice leads to improved performance, that is, higher-order accuracy, in the sample mean case [cf. Singh (1981)]. Nevertheless, in instances where the i.i.d. bootstrap fails, it was observed by many researchers, most notably K. Athreya, J. Bretagnolle, M. Arcones and E. Giné, that bootstrap consistency can be restored by taking a smaller resample size b satisfying (2); see Politis, Romano and Wolf (1999) for references but also note the early paper by Swanepoel (1986).

As a matter of fact, the i.i.d. bootstrap with resample size b immediately inherits the general validity of subsampling in the case where $b = o(\sqrt{n})$ in which case the question with/without replacement is immaterial; see Politis and Romano (1992c, 1993) or Politis, Romano and Wolf (1999, Corollary 2.3.1). However, to relax the condition $b = o(\sqrt{n})$ to $b = o(n)$, some extra structure is required that has to be checked on a case-by-case basis; see Bickel, Götze and van Zwet (1997).

Finally, note that in the case of the sample mean of i.i.d. data (with finite variance), the bootstrap with resample size n outperforms both subsampling and the bootstrap with smaller resample size, as well as the asymptotic normal approximation. Although the performance of subsampling can be boosted by the use of extrapolation and interpolation techniques—see, for example, Booth and Hall (1993), Politis and Romano (1995), Bickel, Götze and van Zwet (1997) and Bertail and Politis (2001)—the understanding has been that subsampling sacrifices some accuracy for its extremely general applicability; this was also the viewpoint adopted in Politis, Romano and Wolf (1999, Chapter 10). Nevertheless, some very recent results of Sakov and Bickel (2000) and Arcones (2001) indicate that this is not always the case: for the sample median, the i.i.d. bootstrap with resample size $b = o(n)$ has improved accuracy as compared to the bootstrap with resample size n; an analogous result is expected to hold for subsampling.

2.2 Block Resampling

In the previous section, we saw the connection of subsampling in the i.i.d. case to the delete-d jackknife. Recall that the standard delete-1 jackknife is commonly attributed to Tukey (1958) and Quenouille (1949, 1956). Interestingly, Quenouille's work was focused on blocking methods in a time series context, and thus it was the precursor of (block) subsampling; of course, blocking methods for estimation in time series go back to Bartlett (1946) as previously mentioned.

After Carlstein's (1986) subseries variance estimator [as well as Hall's (1985) retiling ideas for spatial data] the time was ripe for Künsch's (1989) introduction of the block bootstrap where, instead of recomputing the statistic $\hat{\theta}_b$ on the smaller blocks of the type $B_i = (X_i, \ldots, X_{i+b-1})$, a new bootstrap pseudo-series $X_1^*, \ldots, X_l^*$ is created (with $l = kb \simeq n$) by joining together k $(= [n/b])$ blocks chosen randomly (and with replacement) from the set $\{B_1, \ldots, B_{n-b+1}\}$; then the statistic can be recomputed from the new, full-size pseudo-series $X_1^*, \ldots, X_l^*$.

Thus, 10 years after Efron's (1979) pioneering paper, Künsch's (1989) breakthrough indicated the starting point for the focus of research activity on this new area of block resampling. The closely related work by Liu and Singh (1992), which was available as a preprint in the late 1980's, was also quite influential at that point because of its transparent and easily generalizable approach.

Examples of some early work on block resampling for stationary time series include the blocks-of-blocks bootstrap of Politis and Romano (1992a), which extended the applicability of the block bootstrap to parameters associated with the whole infinite-dimensional probability law of the series; the circular bootstrap of Politis and Romano (1992b), in which the bootstrap distribution is automatically centered correctly; and the stationary bootstrap of Politis and Romano (1994a), which joins together blocks of random length—having a geometric distribution with mean b—and thus generates bootstrap sample paths that are stationary series themselves.

As in the i.i.d. case, the validity of the block-bootstrap methods must be checked on a case-by-case basis, and this is usually facilitated by an underlying asymptotic normality; typically, a condition such as (2) is required. For instance, the block-bootstrap estimator of the distribution of $\sqrt{n}(\bar{X}_n - \mu)$ is essentially consistent only if the latter tends to a Gaussian limit, a result that parallels the i.i.d. situation; see Giné and Zinn (1990) and Radulovic (1996). To see why, note that in the sample mean case the block-bootstrap distribution is really a (rescaled) k-fold convolution of the subsampling distribution $\hat{J}_{b,\mathrm{SUB}}$ with itself; since $k = [n/b] \to \infty$ under (2), it is intuitive that the block-bootstrap distribution has a strong tendency toward Gaussianity.

Interestingly, the block-bootstrap estimator $\hat{\sigma}^2_{b,\mathrm{BB}}$ of $\mathrm{Var}(\sqrt{n}\bar{X}_n)$ is identical to the subsampling estimator $\hat{\sigma}^2_{b,\mathrm{SUB}}$ which in turn is equivalent to the Bartlett estimator $2\pi \hat{f}_{b,\lambda_\mathrm{B}}(0)$; recall that the latter is suboptimal due to its rather large bias of order $O(1/b)$. In addition, the circular and stationary bootstrap estimators of $\mathrm{Var}(\sqrt{n}\bar{X}_n)$ suffer from a bias of the same order of magnitude; see Lahiri (1999). However, if variance estimation is our objective, we can construct a highly accurate estimator via a linear combination of block resampling/subsampling estimators, in the same way that the trapezoidal lag window spectral estimator $\hat{f}_{b,\lambda_\mathrm{T}}$ can be gotten by a linear combination of two Bartlett estimators. The simplest such proposal is to define

$$\tilde{\sigma}^2_b = 2\hat{\sigma}^2_{b,\mathrm{BB}} - \hat{\sigma}^2_{b/2,\mathrm{BB}},$$

where b is now an even integer. The estimator $\tilde{\sigma}^2_b$ has bias (and MSE) that is typically orders of magnitude less than those of $\hat{\sigma}^2_{b,\mathrm{BB}}$; as a matter of fact, $\tilde{\sigma}^2_b \simeq 2\pi \hat{f}_{b,\lambda_\mathrm{T}}(0)$ which, as previously alluded to, enjoys some optimality properties. For example, if the covariance $\gamma(s)$ has an exponential decay—as is the case in stationary ARMA models—then the bias of $\tilde{\sigma}^2_b$ becomes $O(\xi^{-b})$ for some $\xi > 1$; see Politis and Romano (1995), Politis, Romano and Wolf (1999, Section 10.5) or Bertail and Politis (2001).

A theoretical disadvantage of $\tilde{\sigma}^2_b$ is that it is not almost surely nonnegative; this comes as no surprise since the trapezoidal lag window corresponds to a higher-order (actually, *infinite-order*) smoothing kernel: all kernels of order higher than 2 share this problem. Notably, the fast convergence of $\tilde{\sigma}^2_b$ to a nonnegative limit practically alleviates this problem. Nevertheless, the implication is that $\tilde{\sigma}^2_b$ is not associated with some bootstrap distribution for $\bar{X}_n$, as is the case with $\hat{\sigma}^2_{b,\mathrm{BB}}$; if $\tilde{\sigma}^2_b$ could have been gotten as the variance of a probability distribution (empirical or not), then it should be nonnegative by necessity.

To come up with a method that gives improved variance estimation while at the same time being able to produce bootstrap pseudo-series and a bootstrap distribution for $\bar{X}_n$, the following key idea may be employed: the B_i blocks may be *tapered*, that is, their end points are shrunk toward a target value, before being concatenated to form a bootstrap pseudo-series; this is the tapered block-bootstrap method of Paparoditis and Politis (2001a, 2002b) which indeed leads to a variance estimator $\hat{\sigma}^2_{b,\mathrm{TBB}}$ that is more accurate than $\hat{\sigma}^2_{b,\mathrm{BB}}$. In particular, the bias of $\hat{\sigma}^2_{b,\mathrm{TBB}}$ is of order $O(1/b^2)$. Although $\hat{\sigma}^2_{b,\mathrm{BB}}$ is typically less accurate than $\tilde{\sigma}^2_b$, the tapered block-bootstrap constructs an estimator of the whole distribution for $\bar{X}_n$ (not just its variance), and this estimator is more accurate than its (untapered) block-bootstrap analog. Intuitively, this preferential treatment (tapering) of the block edges is analogous to the weighting of the boundary points in constructing $\hat{\mu}_{\mathrm{BLUE}}$ as discussed in the Introduction.

2.3 Block Size Choice and Some Further Issues

Recall that the block size b must be chosen in order to practically implement any of the block resampling/subsampling methods. By the equivalence of $\hat{\sigma}^2_{b,\mathrm{BB}}$ and $\hat{\sigma}^2_{b,\mathrm{SUB}}$ to the Bartlett spectral estimator $2\pi \hat{f}_{b,\lambda_\mathrm{B}}(0)$, it is apparent that choosing b is tantamount to the difficult problem of bandwidth choice in nonparametric smoothing problems. As in the latter problem, here, too, there are two main approaches:

(a) *Cross-validation*. A promising subsampling/cross-validation approach for block size choice proposed by Hall, Horowitz and Jing (1995).

(b) *Plug-in methods*. This approach entails working out an expression for the optimal (with respect to some criterion) value of b and then estimating/plugging in all unknown parameters in that expression.

For the plug-in approach to be viable, it is necessary that the "pilot" estimators to be plugged in are accurate and not plagued by a difficult block/bandwidth choice themselves! Both of those issues can be successfully addressed by using pilot estimators of the trapezoidal lag window type $\hat{f}_{b,\lambda_\mathrm{T}}$; see Paparoditis and Politis (2001a) or Politis and White (2001). Note that choosing b in the estimator $\hat{f}_{b,\lambda_\mathrm{T}}$ (or, equivalently, in $\tilde{\sigma}^2_b$) is relatively straightforward: one may pick $b = 2\hat{m}$, where $\hat{m}$ is the smallest integer such that $\hat{\gamma}(k) \simeq 0$ for $k > \hat{m}$. Here $\hat{\gamma}(k) \simeq 0$ really means $\hat{\gamma}(k)$ not significantly different from 0, that is, an implied hypothesis test; see Politis (2001b) for more details.

Finally, note that all of the aforementioned block resampling/subsampling methods were designed for stationary time series. What happens if the time series is not stationary? Although both subsampling and block resampling will not break down under a mild nonstationarity—see, for example, Politis, Romano and Wolf (1999, Chapter 4)—the following interesting cases may be explicitly addressed:

(i) *Seasonal effects*. If the time series X_t can be decomposed as $X_t = Y_t + W_t$, where Y_t is stationary and W_t is random or deterministic of period d, then

block resampling and subsampling will generally remain valid if b is chosen to be an integer multiple of d. In addition, resampling/subsampling should not be done from the set $\{B_1, B_2, \ldots, B_{n-b+1}\}$ but rather from a set of the type $\{B_1, B_{d+1}, B_{2d+1}, \ldots\}$; that is, full overlap of the blocks is no longer recommended—see Politis (2001a).

(ii) *Local stationarity*. Suppose that the time series has a slowly changing stochastic structure in the sense that the joint probability law of $(X_t, X_{t+1}, \ldots, X_{t+m})$ changes smoothly (and slowly) with t for any m; see, for example, Dahlhaus (1997). Then a "local" block bootstrap may be advisable; here a bootstrap pseudo-series is constructed by a concatenation of k blocks of size b (such that $kb \simeq n$), where the jth block of the resampled series is chosen randomly from a distribution (say, uniform) on all the size-b blocks of consecutive data whose time indices are "close" to those in the original block. A rigorous description of the local block bootstrap is given in Paparoditis and Politis (2002c).

(iii) *Unit root—integrated time series*. A useful model in econometrics is a generalization of the random walk notion given by the *unit root* model under which X_t is not stationary but D_t is, where $D_t = X_t - X_{t-1}$ is the series of first differences—see, for example, Hamilton (1994). In this case, a block bootstrap of the differences D_t can be performed yielding $D_1^*, D_2^*, \ldots$, and a bootstrap pseudo-series for X_t can be constructed by "integrating" the D_t^*, that is, by letting $X_t^* = \sum_{i=1}^{t} D_i^*$. This idea is effectively imposing a sample path "continuity" on the bootstrap realization X_t^*; see Paparoditis and Politis (2001c). On the other hand, if a unit root is just suspected to exist, a bootstrap test of the unit root hypothesis (vs. the alternative of stationarity) may be constructed using similar ideas; see Paparoditis and Politis (2003).

2.4 Why Have Bootstrap Methods Been so Popular?

Finally, it is interesting to recall the two main reasons for the immense success and popularity of the i.i.d. bootstrap of Efron (1979):

(a) The bootstrap was shown to give valid estimates of distribution and standard error in "difficult" situations; a prime example is the median of i.i.d. data for which the (delete-1) jackknife was known to fail, and the asymptotic distribution is quite cumbersome.

(b) In "easy" cases where there exist easy-to-construct alternative distribution estimators, for example, the regular sample mean with its normal large-sample distribution, the (Studentized) bootstrap was shown to outperform those alternatives, that is, to possess "second-order accuracy"—see Hall (1992) for details.

In the case of dependent stationary data, under some regularity conditions, the appropriately standardized/Studentized block bootstrap (in all its variations, including the stationary and circular bootstrap) also possesses a "higher-order accuracy" property when estimating the distribution of the sample mean; see Lahiri (1991, 1999) and Götze and Künsch (1996). [Note that it is not advisable to use a simple block-bootstrap estimate of standard error for the Studentization as it does not converge sufficiently fast; however, one may use an estimator such as $\tilde{\sigma}_b$—see Davison and Hall (1993) or Bertail and Politis (2001).] Nevertheless, the gains in accuracy are not as spectacular as in the i.i.d. case; in addition, estimating the standard error accurately is the dominant factor so that it is somewhat of a luxury to think of capturing higher-order moments.

The bottom line is that practically all time series problems fall under the "difficult" category, and thus the aforementioned block resampling/subsampling methods become invaluable. To give a simple example, consider the problem of estimating the distribution of the lag-p sample autocorrelation $\hat{\rho}(k) = \hat{\gamma}(k)/\hat{\gamma}(0)$, for some fixed k, with the purpose of constructing a confidence interval or hypothesis test concerning $\rho(k)$. It is quite straightforward to show, under usual moment and mixing conditions, that the large-sample distribution of $\sqrt{n}(\hat{\rho}(k) - \rho(k))$ is $N(0, V_k^2)$. Unfortunately, the variance V_k^2 is, in general, intractable involving infinite sums of fourth-order cumulants and so on; see, for example, Romano and Thombs (1996). Nevertheless, block resampling and subsampling effortlessly yield consistent estimators of V_k^2 and the distribution of $\sqrt{n}(\hat{\rho}(k) - \rho(k))$.

One may ask how it had been possible in the pre-bootstrap days to practically address such a basic time series inference question regarding the autocorrelations; the answer is that people had to resort to some more restrictive model assumptions. For example, a celebrated formula for V_k^2 exists involving infinite sums of the (unknown) $\rho(\cdot)$ coefficients but no higher cumulants; this formula is due to Bartlett but is valid only under the assumption that the time series $\{X_t\}$ is *linear*; that is, it satisfies an equation of the type

$$X_t = \mu + \sum_{i=-\infty}^{\infty} \psi_i Z_{t-i}, \tag{4}$$

where $Z_t \sim$ i.i.d. $(0, \tau^2)$ and the ψ_j coefficients are absolutely summable—see Brockwell and Davis (1991). Nonetheless, the class of *nonlinear* time series is vast; for example, the simple time series defined by $Y_t = Z_t Z_{t-1}$ for all t is nonlinear and 1-dependent although uncorrelated. Other examples of nonlinear time series models include the ARCH/GARCH and bilinear models that are of particular prominence in econometric analysis; see, for example, Granger and Andersen (1978), Subba Rao and Gabr (1984), Tong (1990), Bollerslev, Chou and Kroner (1992) or Engle (1995).

To give a concrete example, the $\pm 1.96/\sqrt{n}$ bands that most statistical programs like $S+$ automatically overlay on the correlogram [the plot of $\hat{\rho}(k)$ vs. k] are based on Bartlett's formula. These bands are usually interpreted as giving an implied hypothesis test of the first few $\rho(k)$ being 0 or not, $\rho(1)$ in particular; however, this interpretation is true *only* if one is willing to assume (4). If the time series $\{X_t\}$ is not linear, then Bartlett's formula and the $\pm 1.96/\sqrt{n}$ bands are totally uninformative and may be misleading; resampling and subsampling may well give the only practical solution in this case—see, for example, Romano and Thombs (1996).

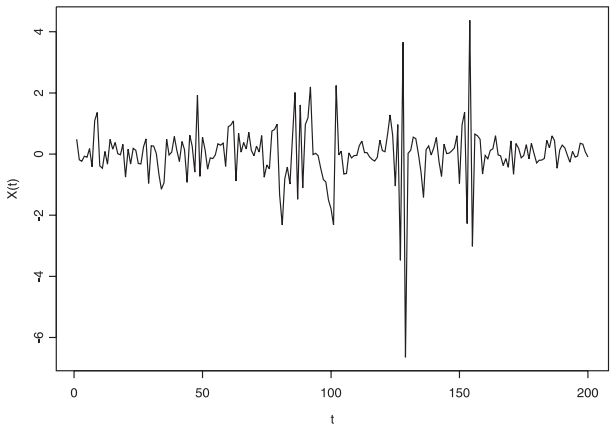
As an illustration, consider a realization $X_1, \ldots, X_{200}$ from the simple ARCH(1) model:

$$X_t = Z_t \sqrt{0.1 + 0.9 X_{t-1}^2},$$

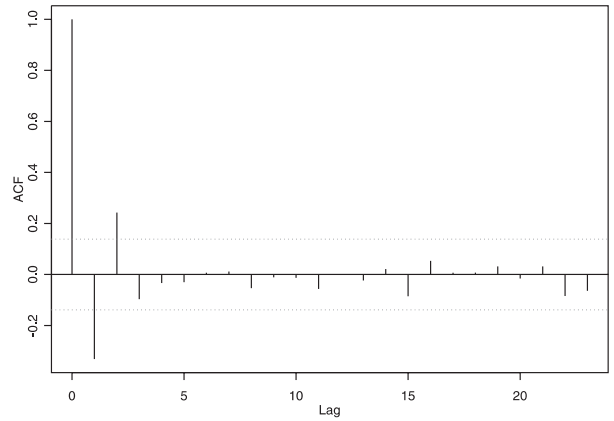
where $Z_t \sim$ i.i.d. $N(0, 1)$; the sample path is plotted in Figure 1(a), while the sample autocorrelation function is given in Figure 1(b).

Note that the time series $\{X_t\}$ is uncorrelated although not independent. Consequently, the $\pm 1.96/\sqrt{n}$ bands are misleading as they indicate that $\hat{\rho}(1)$ is significantly different from 0. Because of the nonlinearity present, the correct 95% confidence limits should be approximately $\pm 4.75/\sqrt{n}$, that is, more than two times *wider* than the ones given by Bartlett's formula. Using the correct confidence limits ± 0.34, the estimate $\hat{\rho}(1) = -0.33$ is seen to *not* be significantly different from 0 (if only barely).

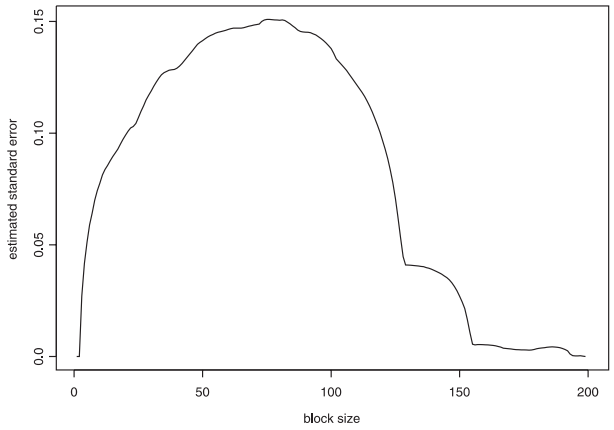
Figure 1(c) depicts the subsampling estimator of the standard error of $\hat{\rho}(1)$ as a function of the subsampling block size b. It is apparent that the block size choice discussed in the previous section is an important issue: for good results, b must be large but small with respect to $n = 200$. With this caveat, the subsampling method seems to be successful in capturing/estimating the correct standard error in this nonlinear setting. Even for a wide range of block size values going from $b = 50$ to 100, the subsampling estimator is 0.14 or above, which is a reasonable approximation of the target value of 0.17.



(a)



(b)



(c)

FIG. 1. (a) *Sample path of an* ARCH(1) *time series with* $n = 200$; (b) *sample autocorrelation of the series with the customary* $\pm 1.96/\sqrt{n}$ *bands*; (c) *subsampling-based estimates of the standard error of* $\hat{\rho}(1)$ *as a function of the subsampling block size* b.

3. MODEL-BASED RESAMPLING FOR TIME SERIES

3.1 Linear Autoregressions

Historically speaking, block resampling/subsampling was not the first approach to bootstrapping non-i.i.d. data. Almost immediately following Efron's (1979) paper, the residual-based bootstrap for linear regression and autoregression was introduced and studied; cf. Freedman (1981, 1984) and Efron and Tibshirani (1986, 1993). In the context of the semiparametric AR(p) model of (1), the residual-based bootstrap amounts to an i.i.d. bootstrap of (a centered version of) the estimated residuals $\hat{Z}_t$ defined by

$$\hat{Z}_t = (X_t - \bar{X}_n) - \hat{\phi}_1(X_{t-1} - \bar{X}_n) - \cdots - \hat{\phi}_p(X_{t-p} - \bar{X}_n),$$

where $\hat{\phi}_j$ are some consistent estimators of ϕ_j, for example, the Yule–Walker estimators. Having constructed bootstrap pseudo-residuals $\hat{Z}_1^*, \hat{Z}_2^*, \ldots$, a bootstrap series $X_1^*, \ldots, X_n^*$ can be generated via the recursion (1) with $\hat{\phi}_j$ in place of ϕ_j; finally, our statistic of interest can be recomputed from the bootstrap series $X_1^*, \ldots, X_n^*$.

If the assumed AR(p) model holds true, then the above residual-based bootstrap works well for the sample mean and other more complicated statistics, and even enjoys a higher-order accuracy property similar to Efron's i.i.d. bootstrap; see Bose (1988). Nevertheless, one never really knows whether the underlying stochastic structure conforms exactly to an AR(p) model, and furthermore one would not know the value of p. A more realistic approach followed by the vast majority of practitioners is to treat (1) only as an approximation to the underlying stochastic structure and to estimate the order p from the data by criteria such as the AIC; see, for example, Shibata (1976). Doing so, one is implicitly assuming that $\{X_t\}$ is a linear time series possessing an AR(∞) representation of the form:

$$(X_t - \mu) = \sum_{i=1}^{\infty} \phi_i (X_{t-i} - \mu) + Z_t, \tag{5}$$

where $Z_t \sim$ i.i.d. $(0, \tau^2)$ and the ϕ_j coefficients are absolutely summable.

Of course, in fitting model (5) to the data, a finite order p must be used which, however, is allowed to increase as the sample size n increases. Thus, the residual-based bootstrap applies *verbatim* to the AR(∞) model of (5) with the understanding that p increases as a function of n; this is the so-called AR(∞) "sieve" bootstrap employed by Swanepoel and van Wyk (1986) and rigorously investigated by Kreiss (1988, 1992), Paparoditis (1992) and Bühlmann (1997), among others. Conditions for the validity of the AR(∞) bootstrap typically require that

$$p \to \infty \quad \text{as } n \to \infty \text{ but with } p^k/n \to 0, \tag{6}$$

where k is a small integer, usually 3 or 4. Consistency still holds when the order p is chosen via the AIC; that is, it is data dependent.

Under assumption (5), Choi and Hall (2000) proved a higher-order accuracy property of the AR(∞) bootstrap in the sample mean case, while Bühlmann (1997, 2002) showed that the AR(∞) bootstrap estimator $\hat{\sigma}^2_{\mathrm{AR}(p)}$ of $\sigma_n^2 = \mathrm{Var}(\sqrt{n}\bar{X}_n)$ has a performance comparable to the aforementioned optimal estimator $\tilde{\sigma}_b^2$. Both $\hat{\sigma}^2_{\mathrm{AR}(p)}$ and $\tilde{\sigma}_b^2$ exhibit an *adaptivity* to the strength of dependence present in the series $\{X_t\}$, that is, they become more accurate when the dependence is weak, and they both achieve a close to $\sqrt{n}$ rate of convergence for series with exponentially decaying covariance $\gamma(s)$, for example, in the case of stationary ARMA models. Not surprisingly, the best performance for $\hat{\sigma}^2_{\mathrm{AR}(p)}$ is achieved when $\{X_t\}$ satisfies a finite-order AR model, that is, a linear Markovian structure, whereas the best performance for $\tilde{\sigma}_b^2$ is achieved when $\{X_t\}$ has an m-dependent structure, for example, a finite-order MA model. Nevertheless, $\tilde{\sigma}_b^2$ is more generally valid. Even among the class of linear time series (4), there are many series that do not satisfy (5), for example, the simple MA(1) model $X_t = Z_t + Z_{t-1}$, where $Z_t \sim$ i.i.d. $(0, \tau^2)$.

3.2 Nonlinear Autoregressions and Markov Processes

Although the class of linear time series is quite rich including, for example, all Gaussian time series, as mentioned before the class of nonlinear time series is vast. It is intuitive that the AR(∞) bootstrap will generally not give good results when applied to data $X_1, \ldots, X_n$ from a nonlinear time series. [Unless, of course, it so happens that the statistic to be bootstrapped has a large-sample distribution that is totally determined by the first two moments of $\{X_t\}$, i.e., by μ and $\gamma(\cdot)$; this is actually what happens in the sample mean case.] See, for example, Bühlmann (2002) for a discussion and some illustrative examples.

The linearity of model (1) is manifested in the fact that the conditional expectation $m(X_{t-1}, \ldots, X_{t-p}) :=$

$E(X_t|X_{t-1},\ldots,X_{t-p})$ is a linear function of the variables $(X_{t-1},\ldots,X_{t-p})$ just as in the Gaussian case. If we do not know and/or assume that $m(\cdot)$ is linear, it may be of interest to estimate its functional form for the purposes of prediction of X_{n+1} given the recent observed past $X_n,\ldots,X_{n-p+1}$. Under some smoothness assumptions on $m(\cdot)$, this estimation can be carried out by different nonparametric methods, for example, kernel smoothing; see, for example, Masry and Tjøstheim (1995). The bootstrap is then called to provide a measure of accuracy of the resulting estimator $\hat{m}(\cdot)$, typically by way of constructing confidence bands. With this objective, many different resampling methods originally designed for nonparametric regression successfully carry over to this autoregressive framework. We mention some prominent examples:

A. *Residual-based bootstrap*. Assuming the structure of a nonlinear (and heteroscedastic) autoregression:

$$X_t = m(X_{t-1},\ldots,X_{t-p}) + v(X_{t-1},\ldots,X_{t-p})Z_t, \tag{7}$$

where $Z_t \sim$ i.i.d. $(0,1)$ and $m(\cdot)$ and $v(\cdot)$ are two (unknown) smooth functions, estimates of the residuals can be computed by $\hat{Z}_t = (X_t - \tilde{m}(X_{t-1},\ldots,X_{t-p}))/\tilde{v}(X_{t-1},\ldots,X_{t-p})$, where $\tilde{m}(\cdot)$ and $\tilde{v}(\cdot)$ are preliminary [oversmoothed as in Härdle and Bowman (1988)] nonparametric estimators of $m(\cdot)$ and $v(\cdot)$. Then an i.i.d. bootstrap can be performed on the (centered) residuals $\hat{Z}_t$, and recursion (7)—with $\hat{m}(\cdot)$ and $\hat{v}(\cdot)$ in place of $m(\cdot)$ and $v(\cdot)$—can be used to generate a bootstrap series $X_1^*,\ldots,X_n^*$ from which our target statistic, for example, $\hat{m}^*(\cdot)$, can be recomputed. The consistency of this procedure under some conditions has been recently shown in a highly technical paper by Franke, Kreiss and Mammen (2002).

B. *Wild and local bootstrap*. Let $\mathbf{X}_{t-1} = (X_{t-1},\ldots,X_{t-p})$ and consider the scatterplot of X_t vs. $\mathbf{X}_{t-1}$; smoothing this scatterplot will give our estimator $\hat{m}(\cdot)$ as well as estimators for higher-order conditional moments if so desired. Interestingly, the wild bootstrap of Wu (1986) and the local bootstrap of Shi (1991) both have a direct application in this dependent context; their respective implementation can be performed exactly as in the independent (nonparametric) regression setting without even having to assume the special structure implied by the nonlinear autoregression (7) which is quite remarkable. Neumann and Kreiss (1998) show the validity of the wild bootstrap in this time series context, while Paparoditis and Politis (2000) show the validity of the local bootstrap under very weak conditions.

Comparing the above two methods, that is, the residual-based bootstrap vs. the wild/local bootstrap, we note the following: the wild/local bootstrap methods are valid under weaker assumptions, as, for example, when (7) is not assumed; on the other hand, the residual-based bootstrap is able to generate new bootstrap stationary pseudo-series $X_1^*,\ldots,X_n^*$ from which a variety of statistics may be recomputed as opposed to just generating new scatterplots to be smoothed.

In order to have our cake and eat it, too, the *Markovian local bootstrap* (MLB for short) was recently introduced by Paparoditis and Politis (2001b, 2002a). To describe it, assume for the sake of argument that the time series $\{X_t\}$ is Markov of order p, with $p=1$ for simplicity. Given the data $X_1,\ldots,X_n$, the MLB procedure constructs a bootstrap sample path $X_1^*,\ldots,X_n^*$ as follows:

(a) Assign X_1^* a value chosen randomly from the set $\{X_1,\ldots,X_n\}$. Suppose this chosen value is equal to X_{t_1} (for some t_1).

(b) Assign X_2^* a value chosen randomly from the set $\{X_{s+1}: |X_{t_1} - X_s| \le h \text{ and } 1 \le s < n-1\}$, where h is a bandwidth-type parameter. Suppose this chosen value is equal to X_{t_2}.

(c) Assign X_3^* a value chosen randomly from the set $\{X_{s+1}: |X_{t_2} - X_s| \le h \text{ and } 1 \le s < n-1\}$. Repeat the above until the full bootstrap sample path $X_1^*,\ldots,X_n^*$ is constructed.

The lag-1 (i.e., $p=1$) MLB described above is capable of capturing the transition probability, as well as the two-dimensional marginal of the Markov process $\{X_t\}$; thus, it can capture the whole infinite-dimensional distribution of $\{X_t\}$ since that is totally determined by the two-dimensional marginal via the Markov property. To appreciate why, consider momentarily the case where $\{X_t\}$ has a finite state space, say $\{1,2,\ldots,d\}$, and suppose that X_{t_1} from part (a) equals the state 1. Then we can even take $h=0$, and the set $\{X_{s+1}: |X_{t_1} - X_s| \le h\}$ is simply the set of all data points whose *predecessor* equaled 1. Choosing randomly from this set is like generating data from the empirical transition function, that is, an estimate of $P(X_{t+1} = \cdot | X_t = 1)$. In the general continuous-state case, consistency of the MLB requires that we let $h \to 0$ as $n \to \infty$, but at a slow enough rate

(e.g., $nh \to \infty$) sufficient to ensure that the cardinality of sets of the type $\{X_{s+1} : |X_t - X_s| \le h\}$ increases with n.

The above MLB algorithm can be easily modified to capture the Markov(p) case for $p \ge 1$. It can then generally be proven—cf. Paparoditis and Politis (2001b, 2002a)—that the MLB bootstrap pseudo-series $X_1^*, X_2^*, \ldots$ is stationary and Markov of order p and that the stochastic structure of $\{X_t^*\}$ accurately mimics the stochastic structure of the original Markov(p) series $\{X_t\}$.

Note that the assumption of a Markov(p) structure is more general than assuming (7). It should be stressed, however, that the applicability of the MLB is *not* limited to Markov processes: the lag-p MLB accurately mimics the $(p+1)$-joint marginal of a *general* stationary $\{X_t\}$ process, and therefore could be used for inference in connection with any statistic whose large-sample distribution only depends on this $(p+1)$-joint marginal; in that sense, the MLB is actually a *model-free* method. Prime examples of such statistics are the aforementioned kernel-smoothed estimators of the conditional moments $m(\cdot)$ and $v(\cdot)$; see, for example, Paparoditis and Politis (2001b).

Finally, note that a closely related method to the MLB is the Markov bootstrap of Rajarshi (1990) that also possesses many favorable properties; see, for example, Horowitz (2003) who also shows a higher-order accuracy property. The Markov bootstrap proceeds by nonparametrically estimating the transition density as a ratio of two kernel-smoothed density estimators (the joint over the marginal). A bootstrap series $X_1^*, \ldots, X_n^*$ is generated by starting at an arbitrary data point and then sampling from this explicitly estimated transition density.

The relation between the MLB and the Markov bootstrap of Rajarshi (1990) is analogous to the relation between Efron's (1979) i.i.d. bootstrap (that samples from the empirical) and the so-called smoothed bootstrap (that samples from a smoothed empirical); for example, the MLB only generates points X_t^* that actually belong to the set of data points $\{X_1, \ldots, X_n\}$, while the Markov bootstrap generates points X_t^* that can be anywhere on the real line. By analogy to the i.i.d. case, it is intuitive here, too, that this extra smoothing may well be superfluous at least in the case of statistics such as the conditional/unconditional moments and functions thereof. Nevertheless, in different situations, for instance, if the statistic under consideration has a large-sample distribution that depends on the underlying marginal densities—as opposed to distributions—this explicit smoothing step may be advisable. An example in the i.i.d. case is given by the sample median (and other nonextreme sample quantiles) where the smoothed bootstrap outperforms the standard bootstrap; see, for example, Hall, DiCiccio and Romano (1989).

ACKNOWLEDGMENTS

Supported in part by NSF Grant DMS-01-04059. Many thanks are due to Professors E. Paparoditis (Cyprus) and J. Romano (Stanford) for constructive comments on this manuscript.

REFERENCES

ARCONES, M. A. (2001). On the asymptotic accuracy of the bootstrap under arbitrary resampling size. *Ann. Inst. Statist. Math.* To appear.

BABU, G. J. and SINGH, K. (1983). Inference on means using the bootstrap. *Ann. Statist.* **11** 999–1003.

BARTLETT, M. S. (1946). On the theoretical specification and sampling properties of autocorrelated time-series. *Suppl. J. Roy. Statist. Soc.* **8** 27–41.

BERTAIL, P. and POLITIS, D. N. (2001). Extrapolation of subsampling distribution estimators: The i.i.d. and strong mixing cases. *Canad. J. Statist.* **29** 667–680.

BICKEL, P. and FREEDMAN, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** 1196–1217.

BICKEL, P., GÖTZE, F. and VAN ZWET, W. R. (1997). Resampling fewer than n observations: Gains, losses, and remedies for losses. *Statist. Sinica* **7** 1–32.

BOLLERSLEV, T., CHOU, R. and KRONER, K. (1992). ARCH modelling in finance: A review of the theory and empirical evidence. *J. Econometrics* **52** 5–59.

BOOTH, J. G. and HALL, P. (1993). An improvement of the jackknife distribution function estimator. *Ann. Statist.* **21** 1476–1485.

BOSE, A. (1988). Edgeworth correction by bootstrap in autoregressions. *Ann. Statist.* **16** 1709–1722.

BROCKWELL, P. and DAVIS, R. (1991). *Time Series*: *Theory and Methods*, 2nd ed. Springer, New York.

BÜHLMANN, P. (1997). Sieve bootstrap for time series. *Bernoulli* **3** 123–148.

BÜHLMANN, P. (2002). Bootstraps for time series. *Statist. Sci.* **17** 52–72.

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

CHOI, E. and HALL, P. (2000). Bootstrap confidence regions computed from autoregressions of arbitrary order. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **62** 461–477.

DAHLHAUS, R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist.* **25** 1–37.

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

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

EFRON, B. and TIBSHIRANI, R. J. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy (with discussion). *Statist. Sci.* **1** 54–77.

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

ENGLE, R., ed. (1995). *ARCH*: *Selected Readings*. Oxford Univ. Press.

FRANKE, J., KREISS, J.-P. and MAMMEN, E. (2002). Bootstrap of kernel smoothing in nonlinear time series. *Bernoulli* **8** 1–37.

FREEDMAN, D. A. (1981). Bootstrapping regression models. *Ann. Statist.* **9** 1218–1228.

FREEDMAN, D. A. (1984). On bootstrapping two-stage least-squares estimates in stationary linear models. *Ann. Statist.* **12** 827–842.

FULLER, W. A. (1996). *Introduction to Statistical Time Series*, 2nd ed. Wiley, New York.

GINÉ, E. and ZINN, J. (1990). Necessary conditions for the bootstrap of the mean. *Ann. Statist.* **17** 684–691.

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

GRANGER, C. and ANDERSEN, A. (1978). *An Introduction to Bilinear Time Series Models*. Vandenhoeck und Ruprecht, Göttingen.

GRENANDER, U. and ROSENBLATT, M. (1957). *Statistical Analysis of Stationary Time Series*. Wiley, New York.

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

HALL, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer, New York.

HALL, P., DICICCIO, T. J. and ROMANO, J. P. (1989). On smoothing and the bootstrap. *Ann. Statist.* **17** 692–704.

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

HAMILTON, J. D. (1994). *Time Series Analysis*. Princeton Univ. Press.

HÄRDLE, W. and BOWMAN, A. (1988). Bootstrapping in nonparametric regression: Local adaptive smoothing and confidence bands. *J. Amer. Statist. Assoc.* **83** 102–110.

HOROWITZ, J. L. (2003). Bootstrap methods for Markov processes. *Econometrica* **71** 1049–1082.

KREISS, J.-P. (1988). Asymptotic statistical inference for a class of stochastic processes. Habilitationsschrift, Faculty of Mathematics, Univ. Hamburg, Germany.

KREISS, J.-P. (1992). Bootstrap procedures for AR(∞) processes. In *Bootstrapping and Related Techniques* (K. H. Jöckel, G. Rothe and W. Sendler, eds.) 107–113. Springer, Berlin.

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

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

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

LIU, R. Y. and SINGH, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring the Limits of Bootstrap* (R. LePage and L. Billard, eds.) 225–248. Wiley, New York.

MASRY, E. and TJØSTHEIM, D. (1995). Nonparametric estimation and identification of nonlinear ARCH time series. *Econometric Theory* **11** 258–289.

NEUMANN, M. and KREISS, J.-P. (1998). Regression-type inference in nonparametric autoregression. *Ann. Statist.* **26** 1570–1613.

PAPARODITIS, E. (1992). Bootstrapping some statistics useful in identifying ARMA models. In *Bootstrapping and Related Techniques* (K. H. Jöckel, G. Rothe and W. Sendler, eds.) 115–119. Springer, Berlin.

PAPARODITIS, E. and POLITIS, D. N. (2000). The local bootstrap for kernel estimators under general dependence conditions. *Ann. Inst. Statist. Math.* **52** 139–159.

PAPARODITIS, E. and POLITIS, D. N. (2001a). Tapered block bootstrap. *Biometrika* **88** 1105–1119.

PAPARODITIS, E. and POLITIS, D. N. (2001b). A Markovian local resampling scheme for nonparametric estimators in time series analysis. *Econometric Theory* **17** 540–566.

PAPARODITIS, E. and POLITIS, D. N. (2001c). The continuous-path block-bootstrap. In *Asymptotics in Statistics and Probability* (M. Puri, ed.) 305–320. VSP Publications, Zeist, The Netherlands.

PAPARODITIS, E. and POLITIS, D. N. (2002a). The local bootstrap for Markov processes. *J. Statist. Plann. Inference* **108** 301–328.

PAPARODITIS, E. and POLITIS, D. N. (2002b). The tapered block bootstrap for general statistics from stationary sequences. *Econom. J.* **5** 131–148.

PAPARODITIS, E. and POLITIS, D. N. (2002c). Local block bootstrap. *C. R. Math. Acad. Sci. Paris.* **335** 959–962.

PAPARODITIS, E. and POLITIS, D. N. (2003). Residual-based block bootstrap for unit root testing. *Econometrica* **71** 813–855.

POLITIS, D. N. (2001a). Resampling time series with seasonal components. In *Frontiers in Data Mining and Bioinformatics*: *Proceedings of the 33rd Symposium on the Interface of Computing Science and Statistics*.

POLITIS, D. N. (2001b). Adaptive bandwidth choice. *J. Nonparametr. Statist.* To appear.

POLITIS, D. N. and ROMANO, J. P. (1992a). 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. (1992b). A circular block-resampling procedure for stationary data. In *Exploring the Limits of Bootstrap* (R. LePage and L. Billard, eds.) 263–270. Wiley, New York.

POLITIS, D. N. and ROMANO, J. P. (1992c). A general theory for large sample confidence regions based on subsamples under minimal assumptions. Technical Report 399, Dept. Statistics, Stanford Univ.

POLITIS, D. N. and ROMANO, J. P. (1993). Estimating the distribution of a Studentized statistic by subsampling. *Bull. Internat. Statist. Inst.* **2** 315–316.

POLITIS, D. N. and ROMANO, J. P. (1994a). The stationary bootstrap. *J. Amer. Statist. Assoc.* **89** 1303–1313.

POLITIS, D. N. and ROMANO, J. P. (1994b). Large sample confidence regions based on subsamples under minimal assumptions. *Ann. Statist.* **22** 2031–2050.

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

POLITIS, D. N., ROMANO, J. P. and WOLF, M. (1999). *Subsampling*. Springer, New York.

POLITIS, D. N. and WHITE, H. (2001). Automatic block-length selection for the dependent bootstrap. *Econometric Rev.* To appear.

QUENOUILLE, M. (1949). Approximate tests of correlation in time-series. *J. Roy. Statist. Soc. Ser. B* **11** 68–84.

QUENOUILLE, M. (1956). Notes on bias in estimation. *Biometrika* **43** 353–360.

RADULOVIC, D. (1996). The bootstrap of the mean for strong mixing sequences under minimal conditions. *Statist. Probab. Lett.* **28** 65–72.

RAJARSHI, M. B. (1990). Bootstrap in Markov sequences based on estimates of transition density. *Ann. Inst. Statist. Math.* **42** 253–268.

ROMANO, J. P. and THOMBS, L. (1996). Inference for autocorrelations under weak assumptions. *J. Amer. Statist. Assoc.* **91** 590–600.

SAKOV, A. and BICKEL, P. (2000). An Edgeworth expansion for the m out of n bootstrapped median. *Statist. Probab. Lett.* **49** 217–223.

SHAO, J. and WU, C.-F. J. (1989). A general theory for jackknife variance estimation. *Ann. Statist.* **17** 1176–1197.

SHERMAN, M. and CARLSTEIN, E. (1996). Replicate histograms. *J. Amer. Statist. Assoc.* **91** 566–576.

SHI, S. G. (1991). Local bootstrap. *Ann. Inst. Statist. Math.* **43** 667–676.

SHIBATA, R. (1976). Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika* **63** 117–126.

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

SUBBA RAO, T. and GABR, M. (1984). *An Introduction to Bispectral Analysis and Bilinear Time Series Models*. *Lecture Notes in Statist.* **24**. Springer, New York.

SWANEPOEL, J. W. H. (1986). A note on proving that the (modified) bootstrap works. *Comm. Statist. Theory Methods*. **15** 3193–3203.

SWANEPOEL, J. W. H. and VAN WYK, J. W. J. (1986). The bootstrap applied to power spectral density function estimation. *Biometrika* **73** 135–141.

TONG, H. (1990). *Non-linear Time Series*: *A Dynamical Systems Approach*. Oxford Univ. Press.

TUKEY, J. W. (1958). Bias and confidence in not-quite large samples (abstract). *Ann. Math. Statist.* **29** 614.

WU, C.-F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis (with discussion). *Ann. Statist.* **14** 1261–1350.