The Dependent Wild Bootstrap

Xiaofeng SHAO

We propose a new resampling procedure, the dependent wild bootstrap, for stationary time series. As a natural extension of the traditional wild bootstrap to time series setting, the dependent wild bootstrap offers a viable alternative to the existing block-based bootstrap methods, whose properties have been extensively studied over the last two decades. Unlike all of the block-based bootstrap methods, the dependent wild bootstrap can be easily extended to irregularly spaced time series with no implementational difficulty. Furthermore, it preserves the favorable bias and mean squared error property of the tapered block bootstrap, which is the state-of-the-art block-based method in terms of asymptotic accuracy of variance estimation and distribution approximation. The consistency of the dependent wild bootstrap in distribution approximation is established under the framework of the smooth function model. In addition, we obtain the bias and variance expansions of the dependent wild bootstrap variance estimator for irregularly spaced time series on a lattice. For irregularly spaced nonlattice time series, we prove the consistency of the dependent wild bootstrap for variance estimation and distribution approximation in the mean case. Simulation studies and an empirical data analysis illustrate the finite-sample performance of the dependent wild bootstrap. Some technical details and tables are included in the online supplemental material.

KEY WORDS: Block bootstrap; Irregularly spaced time series; Lag window estimator; Tapering; Variance estimation.

1. INTRODUCTION

The nonparametric block-based bootstrap for time series has been an active area of research since Künsch (1989) and Liu and Singh (1992) independently introduced the moving block bootstrap (MBB). As an important extension of Efron's iid bootstrap to dependent observations, the MBB can be used to approximate the sampling distributions and variances of many complicated statistics in time series. To capture temporal dependence nonparametrically, the MBB samples the overlapping blocks with replacement and then pastes the resampled blocks together to form a bootstrap sample. Based on the idea of resampling blocks, a few variants of the MBB have been developed, including the nonoverlapping block bootstrap (NBB) (Carlstein 1986), the circular block bootstrap (CBB) (Politis and Romano 1992), the stationary bootstrap (SB) (Politis and Romano 1994), the matched block bootstrap (Carlstein et al. 1998), and the tapered block bootstrap (TBB) (Paparoditis and Politis 2001, 2002), among others. (See Lahiri 2003a for the differences and similarities of these block-based methods.)

In the literature, it seems that the theoretical analysis and applications of the aforementioned block-based methods have been limited mainly to regularly spaced time series. In practice, however, irregularly spaced time series are quite common. As mentioned by Hipel and Mcleod (1994, p. 693), "time series with missing observations or, equivalently, time series where the measurements are taken at unequal time intervals, occur quite often in practice in various fields ... the problem of missing values in data sequences happens frequently in environmental engineering." The irregularity could be due to missing observations in an evenly spaced time series or to the observations taken at randomly sampled time points. There is considerable interest in time series analysis for irregularly spaced data. Statistical methods tailored to regularly spaced time series often need suitable modification to accommodate the irregularity. (See Parzen 1983 for a collection of papers devoted to statistical methodology for irregularly spaced time series.)

The block-based bootstrap methods have been extended to spatial settings for both regular lattice data and irregular nonlattice observations (see, e.g., Politis and Romano 1993; Politis, Paparoditis, and Romano 1999; Lahiri and Zhu 2006; Zhu and Lahiri 2006). Because time series can be considered a one-dimensional analog of spatial data, this might suggest that the aforementioned block-based bootstrap methods can be applied directly to irregularly spaced time series with no difficulty. However, we contend that the use of the block-based bootstrap methods, although theoretically justified, is not convenient for irregular spatial data or regular lattice data in a sampling region of irregular shape. These methods often require a careful partition of the sampling region into incomplete and complete blocks, because the partition usually depends on spatial configuration of the data. This implementational disadvantage, which weakens the automatic feature of bootstrap method, also carries over to irregularly spaced time series.

In this article we propose a new bootstrap method, called the dependent wild bootstrap (DWB), that is generally applicable to stationary, weakly dependent data. The development in this article is confined to time series; extension of the DWB to spatial data will be done elsewhere. The DWB extends the traditional wild bootstrap (Wu 1986) to the time series setting by allowing the auxiliary variables involved in the wild bootstrap to be dependent, so it is capable of mimicking the dependence in the original series. Unlike the block-based bootstrap methods, no partitioning of the data into blocks is involved in the DWB, and an irregular temporal configuration causes no difficulty in implementation. For the smooth function model, the DWB variance estimator reaches the optimal convergence rate in the mean squared error (MSE), which can be achieved only by the TBB among the existing block-based methods. The favorable bias and MSE properties of the TBB variance estimator over its MBB counterpart have been noted and justified by Paparoditis and Politis (2001, 2002) for regularly spaced time

Xiaofeng Shao is Assistant Professor, Department of Statistics, University of Illinois at Urbana–Champaign, Champaign, IL 61820 (E-mail: *xshao@uiuc.edu*). The author is very grateful to an anonymous referee for many constructive comments that led to substantial improvement of the article. The author thanks the editor and an associate editor for providing helpful suggestions, and Jim Miller for providing the data set used in this work. The research was supported in part by National Science Foundation grants DMS-0804937 and DMS-0724752.

Journal of the American Statistical Association
March 2010, Vol. 105, No. 489, Theory and Methods
DOI: 10.1198/jasa.2009.tm08744

series, but a direct extension of the tapering idea to irregularly spaced time series seems difficult. In contrast, the DWB provides a good alternative to the existing block-based bootstrap methods with its superior capability to cope with missing observations or unequally spaced data.

The remainder of the article is organized as follows. Section 2 describes the DWB and its connection to various block-based methods in the context of variance estimation for the sample mean. Section 3 states the consistency of the DWB in distribution approximation for the regularly spaced time series under the framework of a smooth function model. Section 4 presents asymptotic bias and variance expansions for the DWB variance estimator for irregularly spaced time series on a lattice. For time series taken at randomly sampled time points, Section 5 establishes the consistency of the DWB for both variance estimation and distribution approximation in the mean case. The results from simulation studies and an empirical data analysis are reported in Section 6. Section 7 concludes. Technical details are relegated to the Appendix and a supplemental online appendix.

2. THE DEPENDENT WILD BOOTSTRAP

In this section, to facilitate a comparison between the DWB and the block-based methods, we restrict our attention to the regularly spaced univariate time series. The more general setting, which allows for irregularly spaced multivariate time series, is adopted in Section 4. For a stationary time series $\{X_t\}_{t\in\mathbb{Z}}$ with finite variance, let $\mu = \mathbb{E}(X_t)$ and $\gamma_k = \mathrm{cov}(X_0, X_k)$. Given the observations $\mathcal{X}_n = \{X_t\}_{t=1}^n$, we define the DWB pseudo-observations as

$$X_i^* = \bar{X}_n + (X_i - \bar{X}_n)W_i, \qquad i = 1, \dots, n, \tag{1}$$

where $\bar{X}_n = n^{-1}\sum_{t=1}^n X_t$ is the sample mean and $\{W_i\}_{i=1}^n$ are n random variables satisfying the following assumption.

Assumption 2.1. The random variables, $\{W_t\}_{t=1}^n$, are independent of our data, $\mathcal{X}_n$, $\mathbb{E}(W_t) = 0$, $\mathrm{var}(W_t) = 1$ for $t = 1, \dots, n$. Assume that W_t is a stationary process with $\mathrm{cov}(W_t, W_{t'}) = a\{(t - t')/l\}$, where $a(\cdot)$ is a kernel function and $l = l_n$ is a bandwidth parameter. Furthermore, assume that

$$K_a(x) = \int_{-\infty}^{\infty} a(u)e^{-iux}\,du \ge 0 \quad \text{for } x \in \mathbb{R}. \tag{2}$$

The bandwidth parameter, l, plays a similar role as the block size in the block-based methods, as we discuss later. The kernel function $a(\cdot)$ is the same as the lag window function in the definition of the spectral density estimate introduced later [see (3)]. The condition (2) ensures the nonnegative definiteness of the covariance matrix of $\{W_t\}_{t=1}^n$, which can be shown by an elementary argument. Note that (2) is satisfied by a few commonly used kernels, such as Bartlett, Parzen, and quadratic spectral kernels, and it excludes the truncated kernel and Tukey–Hanning kernel (see Andrews 1991, pp. 822–823). The term "dependent wild bootstrap" was coined based on two considerations. On the one hand, it is akin to the wild bootstrap (Wu 1986; Liu 1988; Mammen 1993), which was originally proposed to deal with independent and heteroscedastic errors in the regression problems; on the other hand, unlike the traditional wild bootstrap, the random variables, $\{W_t\}_{t=1}^n$, here are dependent and so are able to capture the dependence in the original sample. Strictly speaking, the series $\{W_t\}_{t=1}^n$ form a triangular array of the type $\{W_{tn} : t = 1, \dots, n; n = 1, 2, \dots\}$. For convenience of presentation, we use $\{W_t\}_{t=1}^n$, and no confusion will arise.

The following notation is used throughout the article. For a random variable ξ, write $\xi \in \mathcal{L}^p$ $(p > 0)$ if $\|\xi\|_p = [\mathbb{E}(|\xi|^p)]^{1/p} < \infty$ and let $\|\cdot\| = \|\cdot\|_2$. Let "$\to_D$" and "$\to_p$" denote convergence in distribution and in probability, respectively, and let $O_p(1)$ and $o_p(1)$ denote being bounded in probability and convergence to 0 in probability, respectively. Let $\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ be a normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Let $\lfloor a \rfloor$ denote the integer part of a and $a \vee b = \max(a, b)$, $a \wedge b = \min(a, b)$ for any $a, b \in \mathbb{R}$. Write $\mathbb{Z}_+ = \{0, 1, 2, \dots\}$. For $\mathbf{v} = (v_1, \dots, v_p)' \in (\mathbb{Z}_+)^p$, $\mathbf{x} \in \mathbb{R}^p$, write $\mathbf{x}^{\mathbf{v}} = \prod_{i=1}^p x_i^{v_i}$, $\mathbf{v}! = \prod_{i=1}^p (v_i!)$. For $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_p)' \in (\mathbb{Z}_+)^p$, let $D^{\boldsymbol{\alpha}}$ denote the differentiable operator $D^{\boldsymbol{\alpha}} = \frac{\partial^{\alpha_1+\cdots+\alpha_p}}{\partial x_1^{\alpha_1}, \dots, \partial x_p^{\alpha_p}}$ on $\mathbb{R}^p$. For a vector $\mathbf{x} = (x_1, \dots, x_p)' \in \mathbb{R}^p$, let $\|\mathbf{x}\|$, $\|\mathbf{x}\|_1 = \sum_{i=1}^p |x_i|$ denote the Euclidean and l^1 norms of $\mathbf{x}$, respectively. The positive constant C is generic and may vary from line to line. All asymptotic statements in the article are with respect to $n \to \infty$ unless specified otherwise.

2.1 Sample Mean

To elucidate the connection between the DWB and other block-based methods, we treat the sample mean case first. Under some moment and weak dependence conditions, we have $T_n := \sqrt{n}(\bar{X}_n - \mu) \to_D \mathrm{N}(0, \sigma_\infty^2)$, where $\sigma_\infty^2 = \sum_{j=-\infty}^{\infty} \gamma_j$ is assumed to be positive. Let $f(\lambda) = (2\pi)^{-1}\sum_{k\in\mathbb{Z}} \gamma_k e^{-ik\lambda}$, $\lambda \in [-\pi, \pi]$ be the spectral density function of the process X_t. Then $\sigma_\infty^2 = 2\pi f(0)$. To construct a confidence interval for μ, one needs to estimate σ_∞^2 or, equivalently, $f(0)$. Let $\hat{\gamma}_k = n^{-1}\sum_{t=|k|+1}^n (X_t - \bar{X}_n)(X_{t-|k|} - \bar{X}_n)$ denote the sample autocovariance at lag k, $|k| \le n - 1$. A large class of spectrum estimators admits the lag-window form

$$\hat{f}_n(\lambda) = (2\pi)^{-1}\sum_{k=1-n}^{n-1} a(k/l)\hat{\gamma}_k \cos(k\lambda), \tag{3}$$

where $a(\cdot)$ is the so-called "lag window function" and $l = l_n$ is the bandwidth. We can take $\hat{\sigma}_\infty^2 = 2\pi\hat{f}_n(0)$. Alternatively, the long run variance σ_∞^2 can be estimated by bootstrap methods. Let P^*, $\mathbb{E}^*$, var^*, cum^* denote the probability, expectation, variance, and cumulant, respectively, conditional on the data $\mathcal{X}_n$. Let $\bar{X}_{n,DWB}^* = n^{-1}\sum_{t=1}^n X_i^*$, with X_i^* as defined in (1). Note that $\mathbb{E}^*(\bar{X}_{n,DWB}^*) = \bar{X}_n$, so the DWB sample has the mean fixed at $\bar{X}_n$. The bootstrapped version of T_n is $T_n^* = \sqrt{n}(\bar{X}_{n,DWB}^* - \bar{X}_n)$, and the DWB estimator of σ_∞^2 is $\hat{\sigma}_{l,DWB}^2 = \mathrm{var}^*(T_n^*)$.

Because later we present the asymptotic equivalence between the DWB variance estimator and its TBB counterpart, here we briefly introduce the TBB procedure (Paparoditis and Politis 2001). Let $w : \mathbb{R} \to [0, 1]$, $w(t) = 0$ if $t \notin [0, 1]$, and $w(t) > 0$ when t is in a neighborhood of $1/2$. Further assume that $w(t)$ is symmetric about $1/2$ and nondecreasing for $t \in [0, 1/2]$. Let $w_n(t) = w\{(t - 0.5)/n\}$, $t = 1, \dots, n$, be the data-tapering sequence. For a fixed block size l, $1 \le l < n$, let $\mathcal{B}_j = \{X_j, X_{j+1}, \dots, X_{j+l-1}\}$ be the jth block, $j = 1, 2, \dots, N = n - l + 1$. The number of blocks in the bootstrap sample is denoted by $b = \lfloor n/l \rfloor$. For the convenience of presentation, we assume that $n = lb$. The TBB consists of two steps: (a) Let $i_0, \ldots, i_{b-1}$ be drawn iid distributed with uniform distribution on the set $\{1, 2, \ldots, N\}$; and (b) for $m = 0, \ldots, b-1$, let

$$X^*_{ml+j} = w_l(j)\frac{l^{1/2}}{\|w_l\|_2}\big(X_{i_m+j-1} - \bar{X}_n\big), \qquad j = 1, \ldots, l, \tag{4}$$

where $\|w_l\|_2 = \{\sum_{t=1}^{l} w_l^2(t)\}^{1/2}$. The TBB estimator of σ^2_∞ is $\hat\sigma^2_{l,TBB} = n\,\mathrm{var}^*(\bar{X}^*_{n,TBB})$, where $\bar{X}^*_{n,TBB}$ is the bootstrapped sample mean based on the TBB pseudoseries $X^*_1, \ldots, X^*_n$ defined in (4). When $w(t) = \mathbf{1}(t \in [0, 1])$ (i.e., no tapering), the TBB boils down to the MBB. Furthermore, if $(i_0, \ldots, i_{b-1})$ are random draws from the set $\{1 + (j-1)l\}_{j=1}^{b}$, then this corresponds to nonoverlapping tapered block bootstrap.

The asymptotic behavior of $\hat\sigma^2_{l,DWB}$ depends largely on the choice of $\{W_t\}_{t=1}^n$. It turns out that with different choices of W_t, the DWB delivers asymptotic variance estimators equivalent to those based on various block bootstrap methods. Under Assumption 2.1 on W_t, we have

$$\begin{aligned}\hat\sigma^2_{l,DWB} &= n^{-1}\sum_{t,t'=1}^{n}(X_t - \bar{X}_n)(X_{t'} - \bar{X}_n)\,\mathrm{cov}^*(W_t, W_{t'})\\ &= n^{-1}\sum_{h=1-n}^{n-1}\sum_{t=1\vee(1-h)}^{n\wedge(n-h)}(X_t - \bar{X}_n)(X_{t+h} - \bar{X}_n)a(h/l)\\ &= 2\pi\hat{f}_n(0);\end{aligned}$$

that is, it is equivalent to the lag window estimator defined in (3). Thus in the sample mean case, the DWB variance estimator depends on W_t through its covariance kernel $a(\cdot)$ and the bandwidth parameter l. Note that the MBB variance estimator is equivalent to $2\pi\hat{f}_n(0)$ when $a(x) = (1 - |x|)\mathbf{1}(|x| \le 1)$ is the Bartlett kernel (see Politis 2003). For the equivalence between the DWB variance estimator and its TBB counterpart, see Remark 2.1. Often in practice, we can take $\{W_t\}_{t=1}^n$ to be multivariate normal with mean 0 and covariance matrix $\boldsymbol\Sigma_l = (\sigma_{ij})_{i,j=1,\ldots,n}$, where $\sigma_{ij} = a\{(i-j)/l\}$. Let $\underline{\mathbf{W}} = (W_1, \ldots, W_n)'$; then $\underline{\mathbf{W}} = \boldsymbol\Sigma_l^{1/2}\mathbf{Z}_n$, where $\mathbf{Z}_n \sim \mathrm{N}(\mathbf{0}, \mathbf{I}_{n\times n})$, with $\mathbf{I}_{n\times n}$ being the $n \times n$ identity matrix. In the implementation of the DWB, $\boldsymbol\Sigma_l^{1/2}$ need be computed only once for each l and given $a(\cdot)$. The computation of $\boldsymbol\Sigma_l^{1/2}$ is fast for small and moderate sample sizes. For large n, more efficient algorithms, such as circular embedding, can be used to generate $\underline{\mathbf{W}}$ (see Dietrich and Newsam 1997). It is worth mentioning that the auxiliary variables $\underline{\mathbf{W}}$ do not have to follow multivariate normal distribution (see Example 4.1).

Let $\tilde{\mathcal{B}}_j = \{j, j+1, \ldots, j+l-1\}$ be the jth block of indexes corresponding to $\mathcal{B}_j$. In general, for the MBB, NBB, CBB, and TBB, due to the mechanism of resampling blocks of fixed size, X^*_i and $X^*_{i'}$ are independent conditional on the data $\mathcal{X}_n$ if the indexes i and i' are not in the same block, $\tilde{\mathcal{B}}_{(j-1)l+1}, j = 1, \ldots, b$. In particular, the dependence between neighboring observations X_l and X_{l+1} are not preserved by its block bootstrap counterpart, whereas for the DWB, dependence of the bootstrap sample is reflected through the dependence of $\{W_t\}_{t=1}^n$. Assuming that the W_t's are l-dependent (which is assumed in our theory) and Assumption 2.1 holds, X^*_i and $X^*_{i'}$ are conditionally dependent (independent) whenever $|i - i'| \le l$ ($|i - i'| > l$). Thus in the DWB, the block structure is enforced on the covariance matrix of the variables $\{W_t\}_{t=1}^n$, but not on the bootstrap sample itself. So, unlike in the block-based methods, the dependence between X_l and X_{l+1} is to some extent preserved by the DWB sample. Furthermore, we note that the DWB sample is no longer stationary conditional on the data.

2.2 Bias and Variance

To state the consistency of $\hat\sigma^2_{l,DWB}$ as an estimator of σ^2_∞, we introduce some assumptions on the lag window function $a(\cdot)$.

Assumption 2.2. Assume that $a : \mathbb{R} \to [0, 1]$ is symmetric and has compact support on $[-1, 1]$, $a(0) = 1$ and $\lim_{x\to 0}\{1 - a(x)\}/|x|^q = k_q \ne 0$ for some $q \in (0, 2]$.

Several commonly used windows (kernels) in spectral analysis, such as Bartlett, Parzen, and Tukey–Hanning windows, satisfy Assumption 2.2. Furthermore, note that $q = 1$ for the Bartlett window and $q = 2$ for the Parzen and Tukey–Hanning windows.

Proposition 2.1. Suppose that Assumption 2.1 on W_t and Assumption 2.2 (with $q = 2$) on $a(\cdot)$ hold. (a) Assume that $X_t \in \mathcal{L}^2$, $\sum_{k=1}^\infty k^2|\gamma_k| < \infty$, and $1/l + l/n^{1/3} = o(1)$. Then

$$\mathbb{E}(\hat\sigma^2_{l,DWB}) = \sigma^2_\infty + \Gamma/l^2 + o(1/l^2), \tag{5}$$

where $\Gamma = -k_2\sum_{k=-\infty}^{\infty}k^2\gamma_k$. (b) Assume that $X_t \in \mathcal{L}^4$, $\sum_{k_1,k_2,k_3\in\mathbb{Z}}|\mathrm{cum}(X_0, X_{k_1}, X_{k_2}, X_{k_3})| < \infty$ and $\sum_{k\in\mathbb{Z}}|\gamma_k| < \infty$. Let $\Delta = 2\sigma^4_\infty\int_{-1}^{1}a^2(x)\,dx$. If $1/l + l/n = o(1)$, then

$$\mathrm{var}(\hat\sigma^2_{l,DWB}) = \Delta \cdot l/n + o(l/n). \tag{6}$$

Proof. The assertions (5) and (6) are special cases of the bias and variance expressions of the lag window estimator at zero frequency. Thus part (a) follows from theorem 9.4.3 of Anderson (1971) (also see Priestley 1981, p. 459), and part (b) holds in view of eqs. (3.9)–(3.12) of Rosenblatt (1984).

Remark 2.1. Let $w * w(t) = \int_{-1}^{1}w(x)w(x+|t|)\,dx$ be the self-convolution of $w(t)$ and $a(x) = w * w(x)/w * w(0)$. Assuming that $w * w(t)$ is twice continuously differentiable at the origin and some other regularity conditions, Paparoditis and Politis (2001) showed that $\mathbb{E}(\hat\sigma^2_{l,TBB}) - \sigma^2_\infty = \Gamma/l^2 + o(1/l^2)$ and $\mathrm{var}(\hat\sigma^2_{l,TBB}) = \Delta \cdot l/n + o(l/n)$. In other words, the DWB variance estimator and the TBB counterpart are asymptotically equivalent provided that $a(x) = w*w(x)/w*w(0)$ and the same bandwidth l is used. Thus the favorable bias and MSE properties of the TBB variance estimator over other block-based counterparts in the mean case automatically carry over to the DWB on choosing $q = 2$.

Fix $q = 2$. Following Paparoditis and Politis (2001), the MSE is minimized when the bandwidth $l_n^{opt} = (4\Gamma^2/\Delta)^{1/5}n^{1/5}$, which gives $\mathrm{MSE}^{opt} = (\Gamma^{2/5}\Delta^{4/5}5 \cdot 4^{-4/5})n^{-4/5}\{1 + o(1)\}$. Given the covariance structure $\{\gamma_k\}_{k\in\mathbb{Z}}$, the lag window that minimizes MSE^{opt} is the one that minimizes the quantity $|a''(0)|\{\int_{-1}^{1}a^2(x)\,dx\}^2$. Paparoditis and Politis (2001) considered the following family of trapezoidal functions: $w_c^{\mathrm{trap}}(t) = (t/c)\mathbf{1}(t \in [0, c)) + \mathbf{1}(t \in [c, 1-c]) + \{(1-t)/c\}\mathbf{1}(t \in (1-c, 1])$. It was found that when $c = 0.43$, the expression $|a''(0)| \times \{\int_{-1}^{1}a^2(x)\,dx\}^2$ reaches its minimum; thus $w_{0.43}^{\mathrm{trap}}$ is used in our simulation studies. (Also see Andrews 1991 for consideration of the optimal kernel in a similar context.)

3. DISTRIBUTION APPROXIMATION VIA THE DEPENDENT WILD BOOTSTRAP

As noted by Paparoditis and Politis (2002), applicability of the block-based bootstrap methods is limited to linear or approximately linear statistics that are root-n consistent and asymptotically normal. For example, suppose that X_t is a univariate time series with marginal distribution F and the quantity of interest is $\theta = T(F)$. A natural estimator of θ is $\hat{\theta}_n = T(\rho_n)$, where $\rho_n = n^{-1}\sum_{t=1}^n \delta_{X_t}$ is the empirical measure, with δ_x representing a unit mass on point x. For an approximately linear statistic $T(\rho_n)$, it admits an expansion in a neighborhood of F, that is, $T(\rho_n) = T(F) + n^{-1}\sum_{t=1}^n \mathrm{IF}(X_t; F) + R_n$, where $\mathrm{IF}(x; F)$ is the influence function (Hampel et al. 1986) defined by $\mathrm{IF}(x; F) = \lim_{\epsilon\downarrow 0}[T\{(1-\epsilon)F + \epsilon\delta_x\} - T(F)]/\epsilon$ and R_n is the remainder term. Under suitable conditions, $n\,\mathrm{var}(\hat{\theta}_n) = n^{-1}\,\mathrm{var}\{\sum_{t=1}^n \mathrm{IF}(X_t; \rho_n)\} + o(1)$.

To see whether the DWB is applicable to the approximately linear statistics, we follow Hall and Mammen (1994) and interpret the DWB in terms of the generation of random measures. Given the sample $\mathcal{X}_n$, the bootstrapped measure ρ_n^* (corresponding to the DWB) can be considered a random distribution with weights at the points $X_1, \ldots, X_n$. Specifically, we can write $\rho_n^* = n^{-1}\sum_{i=1}^n (W_i + 1 - \bar{W}_n)\delta_{X_i}$, where $\bar{W}_n = n^{-1}\sum_{t=1}^n W_t$ and $\{W_t\}_{t=1}^n$ are random variables satisfying Assumption 2.1. In the case where $T(F) = \int x\,dF$, the foregoing formulation amounts to $T(\rho_n) = \bar{X}_n$ and $T(\rho_n^*) = n^{-1}\sum_{t=1}^n (W_t + 1 - \bar{W}_n)X_t = \bar{X}_n + n^{-1}\sum_{t=1}^n W_t(X_t - \bar{X}_n)$, which coincides with the bootstrapped sample mean under the definition in (1). For more general nonlinear statistics, it may be difficult to obtain bootstrap samples, because ρ_n^* is not a valid probability measure. But if $\mathrm{IF}(X_t; \rho_n)$ were known once the data were observed, then the DWB could be applied directly to $\mathrm{IF}(X_t; \rho_n)$. Note that the formulation in the smooth function model is given later. The tapering in the TBB of Paparoditis and Politis (2002) is also applied to $\mathrm{IF}(X_t; \rho_n)$. On the other hand, $\mathrm{IF}(X_t; \rho_n)$ may not be known for some important statistics. For example, when $T(F) = F^{-1}(1/2)$ (i.e., the median of X_1), a natural estimator is $T(\rho_n) =$ sample median. Because $\mathrm{IF}(X_t; F) = [1 - 2\mathbf{1}\{X_t \le F^{-1}(1/2)\}]/[2F'\{F^{-1}(1/2)\}]$ depends on the unknown marginal density function of X_1, $\mathrm{IF}(X_t; \rho_n)$ is unknown in practice. Therefore, the DWB and the TBB of Paparoditis and Politis (2001, 2002) are not directly applicable to variance estimation and distribution approximation in this setting. In contrast, the MBB still works, so the applicability of the DWB is not as wide as that of the MBB for regularly spaced time series.

In the sequel, we consider the class of estimators within the framework of the "smooth function model" (Hall 1992; Lahiri 2003a). This framework is sufficiently general to include many statistics of practical interest, such as autocovariance, autocorrelation, the Yule–Walker estimator, and other interesting statistics in time series. Let $\{\mathbf{X}_t\}_{t\in\mathbb{Z}}$ be a stationary process in $\mathbb{R}^p$ with $\boldsymbol{\mu} = \mathbb{E}(\mathbf{X}_t)$. The quantity of interest is $\theta_0 = H(\boldsymbol{\mu})$, where $H:\mathbb{R}^p \to \mathbb{R}$ is a smooth function. Write $\sigma_n^2 = n\,\mathrm{var}(\hat{\theta}_n)$, where $\hat{\theta}_n = H(\bar{\mathbf{X}}_n)$. Let $\nabla(\mathbf{x}) = \{\partial H(\mathbf{x})/\partial x_1, \partial H(\mathbf{x})/\partial x_2, \ldots, \partial H(\mathbf{x})/\partial x_p\}'$ be the vector of first-order partial derivatives of H at $\mathbf{x}$. Write $\nabla = \nabla(\boldsymbol{\mu})$ and $\boldsymbol{\Sigma}_\infty = \sum_{k=-\infty}^{\infty} \mathrm{cov}(\mathbf{X}_0, \mathbf{X}_k)$. Under some suitable conditions, we have $\sqrt{n}(\hat{\theta}_n - \theta_0) \to_D \mathrm{N}(0, \tau_\infty^2)$, where $\tau_\infty^2 = \nabla'\boldsymbol{\Sigma}_\infty\nabla > 0$. The sampling distribution of $\sqrt{n}\{H(\bar{\mathbf{X}}_n) - H(\boldsymbol{\mu})\}$ can be approximated by its bootstrap analog, $\sqrt{n}\{H(\bar{\mathbf{X}}_{n,DWB}^*) - H(\bar{\mathbf{X}}_n)\}$. Alternatively, because $\mathrm{IF}(\mathbf{X}_t; \rho_n) = \nabla(\bar{\mathbf{X}}_n)'(\mathbf{X}_t - \bar{\mathbf{X}}_n)$, we can apply the DWB to $\mathrm{IF}(\mathbf{X}_t; \rho_n)$, and the resulting bootstrap approximation is $\nabla(\bar{\mathbf{X}}_n)'\sqrt{n}(\bar{\mathbf{X}}_{n,DWB}^* - \bar{\mathbf{X}}_n)$, which is asymptotically equivalent to $\sqrt{n}\{H(\bar{\mathbf{X}}_{n,DWB}^*) - H(\bar{\mathbf{X}}_n)\}$. In practice, when obtaining a closed-form expression for the derivative of $H(\cdot)$ is difficult, $\sqrt{n}\{H(\bar{\mathbf{X}}_{n,DWB}^*) - H(\bar{\mathbf{X}}_n)\}$ is preferred, because it does not involve a calculation of the derivative. We state the consistency only for $\sqrt{n}\{H(\bar{\mathbf{X}}_{n,DWB}^*) - H(\bar{\mathbf{X}}_n)\}$, but the same argument can be applied to show the consistency for $\nabla(\bar{\mathbf{X}}_n)'\sqrt{n}(\bar{\mathbf{X}}_{n,DWB}^* - \bar{\mathbf{X}}_n)$.

Let $\alpha(k)$ denote strong mixing coefficients of the process $\mathbf{X}_t$; by $X_{t,i}$ the ith component of $\mathbf{X}_t$. The following assumptions are needed to state the consistency of the DWB in distribution approximations.

Assumption 3.1. Assume that there exists a $\delta \ge 2$ such that $\sum_{j=1}^\infty \alpha(j)^{\delta/(2+\delta)} < \infty$ and $\mathbb{E}\|\mathbf{X}_1\|^{2+\delta} < \infty$. Also suppose that $\boldsymbol{\Sigma}_\infty$ is nonsingular.

Assumption 3.2. For any $(i_1, i_2, i_3, i_4) \in \{1, 2, \ldots, p\}^4$, we have

$$\sum_{t_1,t_2,t_3=-\infty}^{\infty} \left|\mathrm{cum}\left(X_{0,i_1}, X_{t_1,i_2}, X_{t_2,i_3}, X_{t_3,i_4}\right)\right| < \infty.$$

Theorem 3.1. Assume that the function H is differentiable in a neighborhood of $\boldsymbol{\mu}$, that is, $N_H = \{\mathbf{x} \in \mathbb{R}^p : \|\mathbf{x} - \boldsymbol{\mu}\| \le \epsilon\}$ for some $\epsilon > 0$, $\sum_{|\boldsymbol{\alpha}|=1} |D^{\boldsymbol{\alpha}} H(\boldsymbol{\mu})| \ne 0$, and the first partial derivatives of H satisfy a Lipschitz condition of order $s > 0$ on N_H. Suppose that Assumptions 2.2, 3.1, and 3.2 and $l^{-1} + l/n^{\delta/(2+2\delta)} = o(1)$ hold. Further assume that W_t satisfy Assumption 2.1, $W_t \in \mathcal{L}^{2+\delta}$ and W_t are l-dependent. Then

$$\sup_{x\in\mathbb{R}}\Big|P\big[\sqrt{n}\{H(\bar{\mathbf{X}}_n) - H(\boldsymbol{\mu})\} \le x\big] - P^*\big[\sqrt{n}\{H(\bar{\mathbf{X}}_{n,DWB}^*) - H(\bar{\mathbf{X}}_n)\} \le x\big]\Big| = o_p(1).$$

Note that the l-dependence of W_t implies that for each $t \in \mathbb{N}$, the two sets of random variables $\{W_i, i \le t\}$ and $\{W_i, i \ge t + l + 1\}$ are independent. We impose this assumption to facilitate the blocking argument used in the proof. If the underlying process that generates $\{W_t\}_{t=1}^n$ is Gaussian, then it holds under Assumptions 2.1 and 2.2. The same comment applies to Theorems 4.1, 4.2, and 5.2, where the l-dependence of W_t $[W(t)]$ greatly simplifies the argument in deriving the bias and variance expansions for the DWB variance estimator and proving the consistency of the DWB distribution estimator for irregularly spaced time series.

Remark 3.1. When $\delta > 2$, the main restriction on the bandwidth l allows for $l = O(n^{1/3})$. Thus the optimal bandwidths $l^{opt} = Cn^{1/5}$ for $q = 2$ (see Corollary 4.1) and $l^{opt} = Cn^{1/3}$ for $q = 1$ are both included. The mixing condition in Assumption 3.1 is standard (see Doukhan 1994). The summability of cumulants condition in Assumption 3.2 is commonly adopted in spectral analysis (Brillinger 1975) and is implied by appropriate mixing conditions (Zhurbenko and Zuev 1975; Andrews 1991). In particular, lemma 1 of Andrews (1991) implies that Assumption 3.2 holds if $\mathbb{E}\|\mathbf{X}_1\|^{4+\delta} < \infty$ and $\sum_{j=1}^\infty j^2\alpha(j)^{\delta/(4+\delta)} < \infty$ for some $\delta > 0$.

4. IRREGULARLY SPACED TIME SERIES ON A LATTICE

In practice, the practitioners are often faced with irregularly spaced time series, which arise if there are missing observations in the series. The values missing may be scattered or missing in blocks. Mathematically, suppose that we observe a stationary time series $\mathbf{X}_t \in \mathbb{R}^p$ at time points $t_1 < t_2 < \cdots < t_n$. Again, we are interested in the statistic $\hat{\theta}_n = H(\bar{\mathbf{X}}_n)$, where $\bar{\mathbf{X}}_n = n^{-1} \sum_{j=1}^n \mathbf{X}_{t_j}$. Under some appropriate conditions, theorem 4.3 of Lahiri (2003b) asserts that $n^{-1/2} \sum_{j=1}^n \{\mathbf{X}_{t_j} - \boldsymbol{\mu}\} \to_D \mathrm{N}(0, \boldsymbol{\Sigma}_\infty)$, which implies, in conjunction with the smoothness of $H(\cdot)$, $\sqrt{n}\{H(\bar{\mathbf{X}}_n) - H(\boldsymbol{\mu})\} \to_D \mathrm{N}(0, \tau_\infty^2)$. A problem of interest is to estimate $\tau_n^2 = n \operatorname{var}(\hat{\theta}_n)$ or, equivalently, its limiting variance, $\tau_\infty^2 = \lim_{n\to\infty} \tau_n^2$.

Because the DWB procedure does not involve a partition of the data (or the associated time indexes) into blocks, tapering, and so on, it can be readily applied to irregularly spaced time series. Specifically, the DWB sample is generated by $\mathbf{X}_{t_j}^* = \bar{\mathbf{X}}_n + \{\mathbf{X}_{t_j} - \bar{\mathbf{X}}_n\} W_{t_j}, j = 1, \ldots, n$, where $\{W_{t_j}\}_{j=1}^n$ satisfy the following assumption:

Assumption 4.1. The random variables $\{W_{t_j}\}_{j=1}^n$ are independent of the data and are a realization of a stationary process with $\mathbb{E}\{W_{t_j}\} = 0$, $\operatorname{var}\{W_{t_j}\} = 1$, and $\operatorname{cov}(W_{t_j}, W_{t_{j'}}) = a\{(t_j - t_{j'})/l\}$, where $a(\cdot)$ satisfies (2).

A simple example of $\{W_{t_j}\}_{j=1}^n$ that satisfies Assumption 4.1 is $\mathrm{N}(0, \boldsymbol{\Sigma}_W)$, where $\boldsymbol{\Sigma}_W = [a\{(t_i - t_j)/l_n\}], i, j = 1, \ldots, n$. In principle, the DWB is applicable to any temporal configuration; that is, $\{t_1, \ldots, t_n\}$ can be arbitrary. To facilitate asymptotic analysis, we adopt a one-dimensional analog of the formulation used in most theoretical work for spatial block bootstrap (Lahiri 2003a; Nordman, Lahiri, and Fridley 2007). Let $R_n = \lambda_n R_0$, where λ_n is a sequence of positive real numbers such that $\lambda_n \uparrow \infty$ as $n \to \infty$ and R_0 satisfies the following assumption.

Assumption 4.2. Define R_0 to be a Borel subset of $(-1/2, 1/2]$ containing an open neighborhood of the origin such that for any sequence of positive real numbers $a_n \to 0$, the number of cubes of the scaled lattice $a_n\mathbb{Z}$ that intersect R_0 and R_0^c is $O(1)$ as $n \to \infty$.

Assumption 4.2 is the one-dimensional analog of assumption (1) of Nordman, Lahiri, and Fridley (2007, p. 475) (also see Lahiri and Zhu 2006). It ensures that the effect of the data points lying near the boundary of the line segments is negligible and is satisfied when $R_0 = \bigcup_{i=1}^m (a_i, b_i]$, where $(a_i, b_i] \subset (0, 1]$. We assume that $\{t_1, \ldots, t_n\} = \{t \in \mathbb{Z} \cap R_n\}$; that is, the time points are located on a one-dimensional lattice. The foregoing definition implies that $n \sim \lambda_n |R_0|$, where $|R_0|$ denotes the Lebesgue measure of a set R_0 in $\mathbb{R}$. In the special case where $R_0 = (-1/2, 1/2]$ and $\lambda_n = n$, it reduces to regularly spaced time series. This formulation excludes the case where the observations are taken on randomly sampled time points. The property of the DWB under random sampling is investigated in Section 5.

Denote the bootstrap version of $\hat{\theta}_n$ by $\hat{\theta}_n^* = H(\bar{\mathbf{X}}_n^*)$, with $\bar{\mathbf{X}}_n^* = n^{-1} \sum_{j=1}^n \mathbf{X}_{t_j}^*$, the DWB estimator of τ_∞^2 by $\hat{\tau}_n^2 = n \times \operatorname{var}^*(\hat{\theta}_n^*)$. Note that $\tau_\infty^2 = \sum_{k\in\mathbb{Z}} r_k$, where $r_k = \operatorname{cov}(\boldsymbol{\nabla}'\mathbf{X}_t, \boldsymbol{\nabla}'\mathbf{X}_{t+k})$. To give the (asymptotic) bias and variance expansions of $\hat{\tau}_n^2$, we need to introduce the following conditions.

Condition D_r: $H : \mathbb{R}^p \to \mathbb{R}$ is r-times continuously differentiable and satisfies that $\max\{|D^{\mathbf{v}} H(\mathbf{x})| : |\mathbf{v}| = r\} \le C(1 + \|\mathbf{x}\|^{\kappa_r})$, $\mathbf{x} \in \mathbb{R}^p$ for some integers $\kappa_r \ge 1$, $r = 1, 2, 3$.

Condition M_r: $\mathbb{E}\|\mathbf{X}_1\|^{2r+\delta} < \infty$ and $\Delta(r, \delta) = \sum_{n=1}^\infty n^{2r-1} \times \alpha(n)^{\delta/(2r+\delta)} < \infty$ for some $\delta > 0$.

Condition C_r: For any $(i_1, \ldots, i_s) \in \{1, \ldots, p\}^s$, $2 \le s \le r$,

$$\sum_{t_1, t_2, \ldots, t_{s-1} \in \mathbb{Z}} \left|\operatorname{cum}\left(X_{0,i_1}, X_{t_1,i_2}, \ldots, X_{t_{s-1},i_s}\right)\right| < \infty.$$

Theorem 4.1. Suppose that Assumptions 2.2 (with $q = 2$), 4.1, and 4.2 and Conditions C_8 and D_3 hold, and that M_r holds with $r = (3 + \kappa_3) \vee (2\kappa_2) \vee 4$. Assume that $1/l_n + l_n/n^{1/4} = o(1)$, $\sum_{k\in\mathbb{Z}} |kr_k| < \infty$, and $W_{t_j}, j = 1, \ldots, n$, are l_n-dependent. Then $\mathbb{E}(\hat{\tau}_n^2) = \tau_\infty^2 + B_0 l_n^{-2} + o(l_n^{-2})$, where $B_0 = -k_2 \sum_{k=-\infty}^\infty k^2 r_k$.

Remark 4.1. Lahiri (1999) used conditions D_r and M_r to derive the bias and variance expansions of the block-based bootstrap variance estimator. The constraint on the bandwidth l_n is a bit restrictive because of some technical requirements, but it does not exclude the optimal bandwidth, which is of order $O(n^{1/5})$ (see Corollary 4.1). Theorem 4.1 still holds if we replace τ_∞^2 by τ_n^2, because $\tau_n^2 = \tau_\infty^2 + O(n^{-1/2})$ (cf. lemma 10.1 in Nordman and Lahiri 2004).

Theorem 4.2. Suppose that Assumptions 2.2, 4.1, and 4.2 and Conditions C_{16} and D_3 hold, and that M_r holds with $r = (6 + 2\kappa_3) \vee (4\kappa_2) \vee 8$. Assume that $1/l_n + l_n/n^{1/2} = o(1)$, $\sum_{k\in\mathbb{Z}} |kr_k| < \infty$, and $W_{t_j}, j = 1, \ldots, n$, are l_n-dependent. Then we have $\operatorname{var}(\hat{\tau}_n^2) = D_0 l_n/n\{1 + o(1)\}$, where $D_0 = 2\tau_\infty^4 \int_{-1}^1 a^2(x)\,dx$.

Corollary 4.1. Under the combined assumptions of Theorems 4.1 and 4.2, the optimal bandwidth that minimizes the MSE of the DWB variance estimator is given by $l_n^{opt} = (4B_0^2/D_0)^{1/5} n^{1/5}$, and the corresponding MSE is $\{(4^{-4/5} + 4^{1/5}) B_0^{2/5} D_0^{4/5}\} n^{-4/5}\{1 + o(1)\}$.

Unlike the choice of the block size, the bandwidth l in the DWB does not have to be an integer in practice, so the optimal theoretical MSE can be reached by choosing a noninteger bandwidth. For spatial data on a regular grid, Nordman, Lahiri, and Fridley (2007) obtained the asymptotic bias and variance expansions of the spatial block bootstrap variance estimator under the framework of smooth function model. Their results, when reduced to the one-dimensional case, amount to $1/l_n$ and l_n/n for the order of the leading terms in the bias and variance expansions. Consequently, the order of the leading term in the minimized MSE is $n^{-2/3}$ for both MBB and NBB. Therefore, in terms of the optimal MSE, the MBB (NBB) is inferior to the DWB for both regularly and irregularly spaced time series on a lattice. It is also worth noting that all the theoretical results still hold if we approximate the sampling distribution (or the variance) of $\sqrt{n}\{H(\bar{\mathbf{X}}_n) - H(\boldsymbol{\mu})\}$ by its bootstrap counterpart based on $\boldsymbol{\nabla}(\bar{\mathbf{X}}_n)'\sqrt{n}(\bar{\mathbf{X}}_n^* - \bar{\mathbf{X}}_n)$. In this case, the strong moment conditions in Theorems 4.1 and 4.2 can be relaxed in view of the proof of Theorem 4.1.

Up to this point, we have justified the consistency of the DWB for distribution approximation for regularly spaced time series, and obtained the bias and variance expansions of the DWB variance estimator for irregularly spaced time series on a lattice under the framework of smooth function model. A natural next step is to investigate whether the DWB can offer the second-order correctness, that is, better than normal approximation. In the regularly spaced case, the second-order correctness for the MBB has been studied by Lahiri (1992, 2007) and Götze and Künsch (1996). We have not been able to obtain any results for the DWB, but it is certain that the choice of multivariate normal distribution for W_t, although convenient, would not lead to the second-order correctness in general. If the W_t are multivariate normal, then $\mathrm{cum}(W_i, W_j, W_k) = 0$ for any (i, j, k), which implies that $\mathrm{cum}^*(X_i^*, X_j^*, X_k^*) = 0$ for any $i, j, k = 1, \ldots, n$. So the DWB sample would not be able to match possibly nonzero third cumulants of the original univariate process and would not yield the second-order correctness. If the time series $\mathbf{X}_t$ is Gaussian, we conjecture that the DWB using multivariate normal W_t would yield the second-order correctness for the properly studentized statistic under the framework of smooth function model, although a rigorous proof is well beyond the scope of this article. On the other hand, the $\{W_t\}_{t=1}^n$ do not have to follow a multivariate normal distribution, as illustrated by the following example.

Example 4.1. Let $W_t = (\widetilde{W}_t^2 - 1)/\sqrt{2}$, where $\{\widetilde{W}_t\}_{t=1}^n$ are multivariate normal random variables with mean 0 and covariance matrix $\widetilde{\boldsymbol{\Sigma}}_l = (\tilde{\sigma}_{ij})_{i,j=1,\ldots,n}$, $\tilde{\sigma}_{ij} = \tilde{a}(|i-j|/l)$, where $\tilde{a}(\cdot)$ satisfies (2) and Assumption 2.2. Then it is easy to see that W_t satisfies Assumption 2.1 with $\mathrm{cov}(W_t, W_s) = \mathrm{cov}^2(\widetilde{W}_t, \widetilde{W}_s) = \tilde{a}^2(|t-s|/l)$, that is, $a(x) = \tilde{a}^2(x)$ under our notation. Unlike the multivariate normal case, the marginal distribution of W_t is no longer normal and

$$\begin{aligned}
&\mathrm{cum}(W_0, W_t, W_s)\\
&\quad = 2^{-3/2}\,\mathrm{cum}(\widetilde{W}_0^2, \widetilde{W}_t^2, \widetilde{W}_s^2)\\
&\quad = 2\sqrt{2}\,\mathrm{cov}(\widetilde{W}_0, \widetilde{W}_t)\,\mathrm{cov}(\widetilde{W}_0, \widetilde{W}_s)\,\mathrm{cov}(\widetilde{W}_t, \widetilde{W}_s)\\
&\quad = 2\sqrt{2}\tilde{a}(t/l)\tilde{a}(s/l)\tilde{a}\{(t-s)/l\}\\
&\quad = \sqrt{8a(t/l)a(s/l)a\{(t-s)/l\}} \not\equiv 0.
\end{aligned}$$

In general, given a time series at hand, it is a hard task to design the joint distribution for W_t to match the higher-order cumulants of the unknown data-generating process. Thus we have no practical recommendations as to the joint distribution of W_t if the goal is to achieve the second-order accuracy. In practice, the choice of the kernel $a(\cdot)$ and the bandwidth l, which affects the first-order accuracy, may be more important than the choice of possibly nonnormal joint distribution for W_t [for given l and $a(\cdot)$], which affects only the second-order accuracy. The DWB is especially useful in the case of irregular spaced time series, for which the second-order accuracy of the MBB is not known. Thus it seems fair to regard the DWB as a viable alternative to the block-based methods despite its possible lack of second-order accuracy in the non-Gaussian case.

5. TIME SERIES WITH STOCHASTIC SAMPLING DESIGN

To allow a nonlattice configuration for the time points $\{t_j\}_{j=1}^n$, we adopt the framework of stochastic sampling design, which was used by Lahiri and Zhu (2006) to study the consistency of the spatial block bootstrap for irregularly spaced spatial data. (See Lahiri 2003b and Lahiri and Mukherjee 2004 for earlier work within the same framework.) Because we deal only with time series in this article, we restrict our attention to the one-dimensional case. Assume that $t_j = \lambda_n z_j$, $j = 1, \ldots, n$, where z_j takes values in R_0 and $\{z_j\}_{j=1}^n$ are a realization of the iid random variables $Z_1, \ldots, Z_n$. As described by Lahiri (2003b), this formulation of stochastic design allows a nonuniform density across the region, so the expected number of points in two regions of the same size could be different. In addition, depending on the magnitude of $\kappa := \lim_{n\to\infty} n/\lambda_n$, this formulation accommodates both pure increasing-domain asymptotics (i.e., $\kappa < \infty$) and mixed increasing-domain asymptotics (i.e., $\kappa = +\infty$).

In this section we assume that there is an underlying continuous-time stationary process $\{X(t), t \in \mathbb{R}\}$ and the observations are $\{X(t_j)\}_{j=1}^n$. The notation differs from the $\{X_t, t \in \mathbb{Z}\}$ used in previous sections for a discrete-time process. Correspondingly, we use $\gamma(z_1) = \mathrm{cov}\{X(0), X(z_1)\}$, $C_4(z_1, z_2, z_3) = \mathrm{cum}\{X(0), X(z_1), X(z_2), X(z_3)\}$ to denote the autocovariance and the fourth-order cumulant for $z_1, z_2, z_3 \in \mathbb{R}$. For simplicity, we focus on the mean case; see Remark 5.1 for a discussion of the smooth function model. Let $\mu = \mathbb{E}\{X(t)\}$ and $\bar{X}_n = n^{-1}\sum_{j=1}^n X(t_j)$. We are interested in estimating the distribution and the variance of $\bar{X}_n - \mu$ using the DWB method. Write $\xi_n = \mathrm{var}(\bar{X}_n)$. The DWB sample is obtained as $X^*(t_j) = \bar{X}_n + \{X(t_j) - \bar{X}_n\}W(t_j)$, $j = 1, \ldots, n$, where $\{W(t_j)\}_{j=1}^n$ represents a realization from a continuous-time stationary process $W(t)$, $t \in \mathbb{R}$. Without loss of generality, we assume that $\{Z_n\}_{n\ge 1}$, $\{X(t), t \in \mathbb{R}\}$ and that the bootstrap variables $\{W(t), t \in \mathbb{R}\}$ are all defined on a common probability space $(\Omega, \mathcal{F}, P)$. Let P_Z denote the joint probability distribution of the sequence of iid random variables $Z_1, Z_2, \ldots$ with density $\eta(z)$, $z \in R_0$. We use $\mathbb{E}_Z$ (var_Z) to denote the expectation (variance) with respect to the joint distribution P_Z and use $\mathbb{E}_{X|Z}$ ($\mathrm{var}_{X|Z}$) to denote the conditional expectation (variance) with respect to P_X (i.e., the joint probability distribution for $\{X(t), t \in \mathbb{R}\}$) given $\{Z_n\}_{n\ge 1}$. The following assumption on $\eta(\cdot)$ is assumed throughout this section.

Assumption 5.1. The pdf $\eta(x)$ is continuous, everywhere positive with support $\bar{R}_0$ and $\int_{s\in R_0} \eta(s)\,ds = 1$.

Write $\iota = \int_{s\in R_0} \eta^2(s)\,ds$. Lemma 5.2 of Lahiri (2003b) implies that under appropriate conditions, we have that (a) if $\kappa \in (0, \infty)$, then $n\xi_n \to \gamma(0) + \kappa\iota\int_{\mathbb{R}} \gamma(s)\,ds$, a.s. ($P_Z$) and (b) if $\kappa = \infty$, then $\lambda_n \xi_n \to \iota\int_{\mathbb{R}} \gamma(s)\,ds$, a.s. ($P_Z$). Here a.s. ($P_Z$) means that the result holds with probability 1 under P_Z, that is, for almost all realizations of the sequence $\{Z_n\}$. Lahiri (2003b) regarded the distribution of $\bar{X}_n$ as a conditional distribution given $\{Z_n\}_{n\ge 1}$ and ξ_n as a function of the randomly sampled locations. Whereas in our treatment, we view ξ_n as an unknown quantity, where the randomness due to $\{Z_n\}_{n\ge 1}$ has been removed by the expectation.

Let $\bar{X}_n^* = n^{-1}\sum_{j=1}^n X^*(t_j)$. Let $\hat{\xi}_n = \mathrm{var}^*(\bar{X}_n^*)$ denote the DWB variance estimator of ξ_n. The following theorem shows the consistency of $\hat{\xi}_n$ as an estimator of ξ_n.

Theorem 5.1. Suppose that Assumption 2.2 on $a(\cdot)$, Assumption 4.2 on R_0, and Assumption 5.1 on $\eta(\cdot)$ hold. Assume that $\{W(t_j)\}_{j=1}^n$ satisfies Assumption 4.1, with W_t replaced by $W(t)$. Further assume that $l_n/\sqrt{n} + l_n/\lambda_n = o(1)$,

$$\int_{\mathbb{R}} |\gamma(z)|\, dz < \infty \qquad \text{and} \tag{7}$$

$$\int_{\mathbb{R}^3} |C_4(z_1, z_2, z_3)|\, dz_1\, dz_2\, dz_3 < \infty. \tag{8}$$

We have that (a) if $\kappa \in (0, \infty)$, then $n\hat{\xi}_n \to_p \gamma(0) + \kappa\iota \times \int_{\mathbb{R}} \gamma(s)\, ds$, and (b) if $\kappa = \infty$, then $\lambda_n \hat{\xi}_n \to_p \iota \int_{\mathbb{R}} \gamma(s)\, ds$.

Remark 5.1. It is possible to prove the consistency in the case of smooth function model under additional regularity conditions. In particular, we need to impose certain smoothness conditions on the smooth function (including growth and boundedness conditions on suitable derivatives of the smooth function) and may require stronger moment and weakly dependence conditions on $X(t)$. The argument is expected to resemble that in the proofs of Theorems 4.1 and 5.1, but is more involved.

It is worth noting that the process $W(t)$ should be written as $W_n(t)$, where the dependence on n is suppressed for notational convenience. The following theorem states the consistency of the DWB in terms of distribution approximation.

Theorem 5.2. Suppose that the assumptions in Theorem 5.1 hold. Further suppose that the $\{W(t)\}$ are l-dependent and $W(t) \in \mathcal{L}^4$. Then we have that for case (a) (i.e., $\kappa < \infty$),

$$\sup_{x \in \mathbb{R}} \left| P[\sqrt{n}\{\bar{X}_n - \mu\} \le x] - P^*[\sqrt{n}(\bar{X}_n^* - \bar{X}_n) \le x] \right| = o_p(1),$$

and for case (b) (i.e., $\kappa = \infty$),

$$\sup_{x \in \mathbb{R}} \left| P[\sqrt{\lambda_n}\{\bar{X}_n - \mu\} \le x] - P^*[\sqrt{\lambda_n}(\bar{X}_n^* - \bar{X}_n) \le x] \right| = o_p(1).$$

It is interesting to see that the DWB is consistent under both pure increasing-domain asymptotics and mixed increasing-domain asymptotics. Under the same sampling design, Lahiri and Zhu (2006) showed the consistency for the grid-based block bootstrap in the spatial setting in the sense that it can consistently approximate the sampling distribution of the M-estimator in spatial regression models. In particular, their theorem 2 states the bootstrap consistency in the sense of $P_{\cdot|Z}$-probability a.s. (P_Z). This type of convergence is slightly stronger than what we have in Theorem 5.2. It is certainly an interesting topic to extend the validity of the DWB to their setup with the stronger convergence result. We leave this for future investigation. A comparison between the DWB and the grid-based block bootstrap is made through simulations; see Section 6.2.

6. NUMERICAL EXAMPLES

In previous sections, we discussed the DWB from a theoretical perspective. To corroborate our theoretical findings, here we compare via simulations the finite-sample performance of the DWB and the block-based bootstrap methods for time series with or without missing values in Section 6.1 and for time series with randomly sampled time points in Section 6.2. We provide an empirical illustration in Section 6.3.

6.1 Time Series on a Lattice

Here we focus on the inference of the population mean of a time series. We revisit the following nonlinear autoregressive model used by Paparoditis and Politis (2001):

$$X_t = 0.6\sin(X_{t-1}) + Z_t, \qquad t \in \mathbb{Z}, \tag{9}$$

where $\{Z_t\}$ are iid N(0, 1). Let $n = 200$. Among the block-based bootstrap methods, the theoretical advantage of the TBB over the MBB has been confirmed for (9) through simulation studies by Paparoditis and Politis (2001). For this reason, we only compare the DWB with the TBB in this case. To make the comparison fair, we use $w(x) = w_{0.43}^{\text{trap}}(x)$ and $a(x) = w_{0.43}^{\text{trap}} * w_{0.43}^{\text{trap}}(x)/w_{0.43}^{\text{trap}} * w_{0.43}^{\text{trap}}(0)$ in the TBB and DWB, respectively, so the TBB and DWB variance estimators have the same (asymptotic) MSEs for a given bandwidth l.

For each simulated time series (with no missing values) and each block size, 1000 TBB and DWB pseudoseries are generated to perform variance estimation and distribution approximation. We repeat this procedure 3000 times and plot the empirical MSE and empirical coverage of nominal 95% symmetric confidence intervals as a function of block size in Figure 1(a) and (d). Because our theoretical discussion focuses on the bias and variance of the DWB variance estimator, it seems natural to plot the MSE and compare the optimal bandwidths that minimize the empirical MSE. In light of the fact that MSE favors underestimation in the context of variance estimation, we also plot the ratio of the averaged variance estimates (over 3000 replications) to the true variance $\sigma_n^2 = n\,\text{var}(\bar{X}_n)$ for $n = 200$ and the coefficient of variation for 3000 independent variance estimates in Figure 1(b) and (c). Here σ_n^2 is obtained by Monte Carlo approximations with $n = 200$ and 10^6 replications. It can be seen that the TBB and DWB have comparable empirical performance at a range of block sizes, although in this case the TBB is noticeably inferior to the DWB when the block size exceeds 20. The optimal block sizes for the TBB and DWB are very close in terms of both MSE and coverage. Moreover, the optimal MSEs (i.e., the MSE corresponding to the empirically optimal block size) for the TBB and DWB are almost the same. Figure 1(b) shows the tendency toward underestimation of both the TBB and DWB estimators. The optimal bandwidth that minimizes the distance between the (empirical) ratio and 1 is larger than the optimal bandwidth that minimizes the MSE, which can be explained by the monotonically increasing pattern of the coefficient of variation with respect to the block size [see Figure 1(c)].

To investigate the performance of the DWB when there are missing values, we artificially assume that the observations from the model (9) are missing at the following 10 randomly generated time points: $t = 4, 73, 121, 126, 130, 139, 144, 160, 163, 191$. Thus the effective sample size is $n = 190$. When a time series has a small number of missing values, the MBB (TBB) is expected to work in theory. In practice, we have a few possible choices for dealing with missing values, including the following schemes:

1. Ignore missing values and perform the MBB (TBB). In other words, we need to keep track of the missing patterns in the resampled blocks and use only the nonmissing resampled data. We designate this scheme MBB-I (TBB-I).

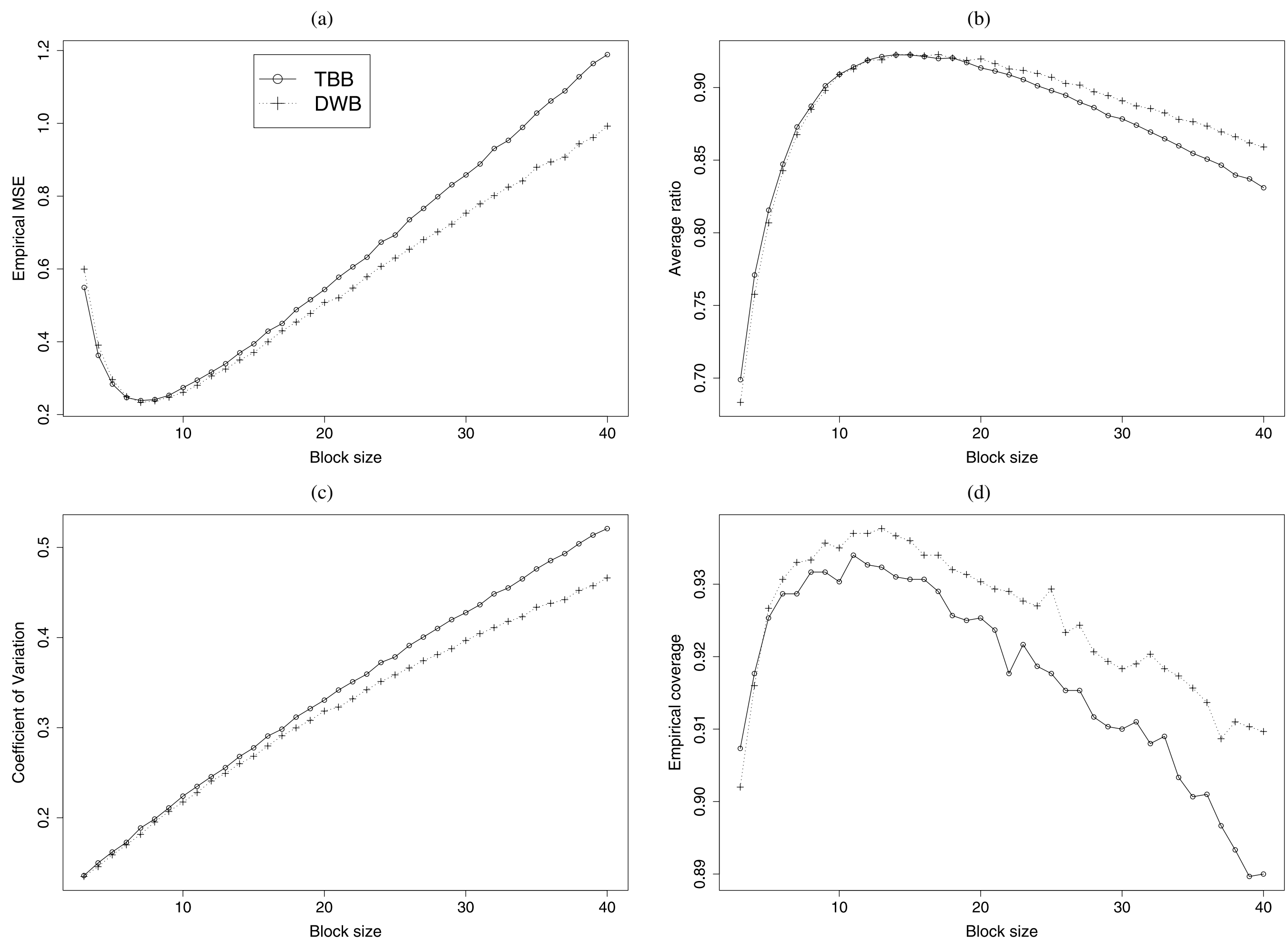

Figure 1. Time series generated from the model (9) with no missing values; sample size, $n = 200$. (a) Empirical MSEs of the TBB and DWB variance estimators of σ_n^2. The largest standard error is 0.034. (b) Ratio of the averaged TBB (DWB) variance estimates (over 3000 replications) to σ_n^2. The largest standard error is 0.070. (c) Coefficient of variation for 3000 independent TBB (DWB) estimates of σ_n^2. The largest standard error is 0.0084, which is calculated using the jackknife method. (d) The empirical coverage probability of a two-sided 95% confidence interval of the mean. The largest standard error is 0.0057.

2. Impute the missing values using the sample mean and then perform the MBB (TBB) on the imputed data set. We designate this scheme MBB-II (TBB-II). Note that more complex imputation methods could be tried, but using sample mean as the imputed value seems simple and natural.
3. Take the missing pattern into account in the block bootstrapping, as described by Nordman, Lahiri, and Fridley (2007). They used a modified MBB scheme (designated MBB-III) in the spatial setting, as we discuss in the next paragraph. To make a fair comparison of the MBB, TBB, and DWB, we let $a(\cdot)$ denote the Bartlett kernel in the DWB procedure (designated DWB1), which can be shown to deliver an asymptotically equivalent variance estimator as the MBB counterpart for regularly spaced time series. We also incorporate the DWB with $a(\cdot) = w_{0.43}^{\text{trap}} * w_{0.43}^{\text{trap}}(x)/w_{0.43}^{\text{trap}} * w_{0.43}^{\text{trap}}(0)$ (designated DWB2) to examine the possible advantage of tapering.

The basic idea of the modified MBB scheme of Nordman, Lahiri, and Fridley (2007) is that for a prespecified block size l, one finds all of the time points (or spatial locations) that contain all of the nonoverlapping complete blocks of size l, and fills in with the corresponding resampled values. Note that the resampled data are only from all of the complete (overlapping) blocks. In general the modified MBB sample has a shorter length due to the ignorance of the boundary time points that are close to the locations of missing values. For large l, the difference between the bootstrap sample size and the original sample size tends to be nonnegligible, which may contribute to the inaccuracy of the bootstrap estimator. In addition, as l increases, the number of candidate blocks becomes less and the number of data points used in resampling decreases. This may lead to a loss of efficiency. These practical disadvantages are confirmed in our simulation results presented later. Note that the TBB of Paparoditis and Politis (2001, 2002) was introduced for the regularly spaced time series, and its extension to accommodate the irregularity (e.g., missing values) is not available in the literature. For time series on a lattice, one can presumably carry out such an extension by combining the modified MBB with the tapering scheme of Paparoditis and Politis (2001, 2002). Here we do not present the results for this modified TBB, but would

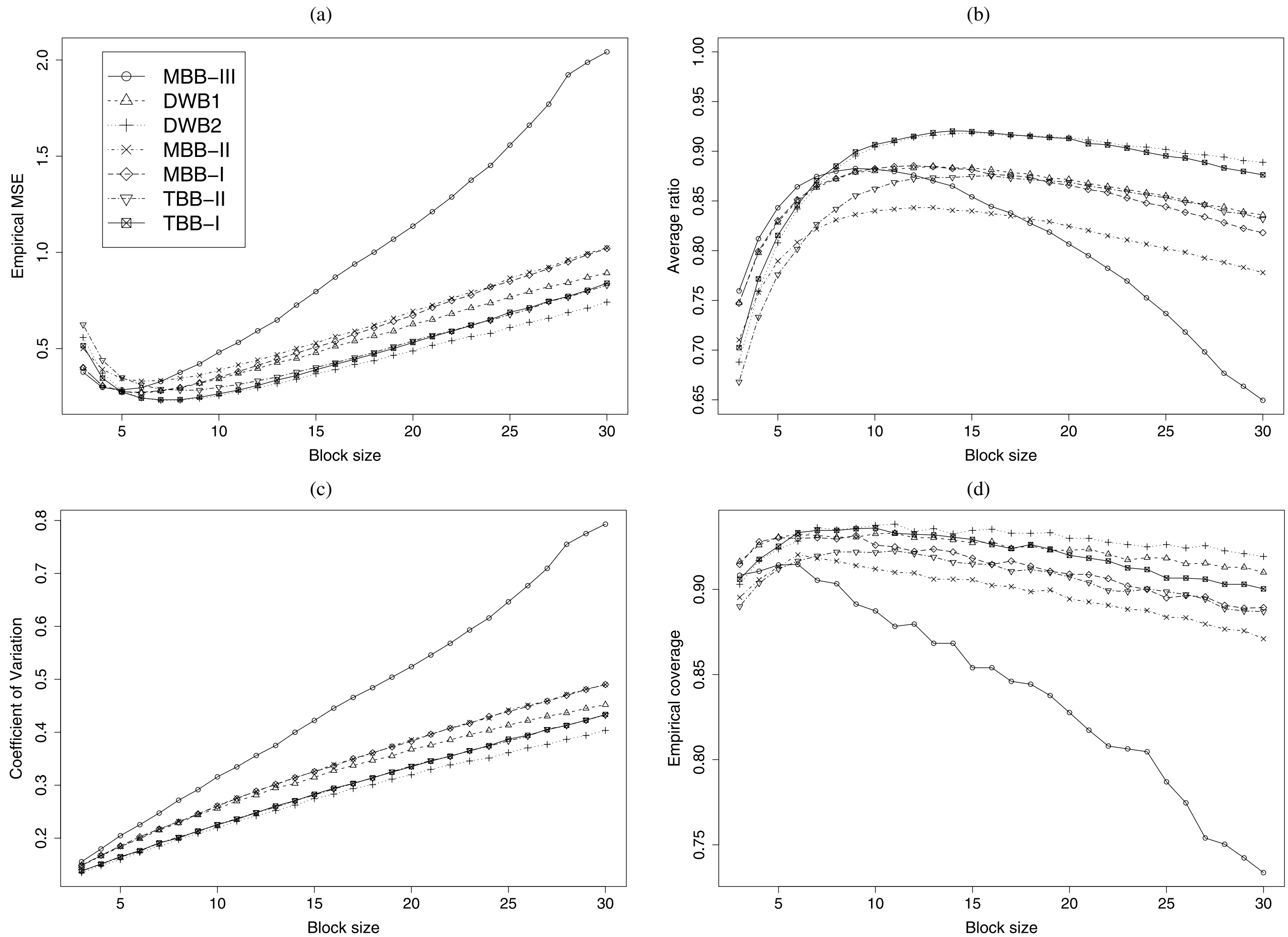

Figure 2. Time series generated from the model (9) with 10 missing values; sample size $n = 190$. The comparison is among MBB-I, MBB-II, MBB-III, DWB1, DWB2, TBB-I, and TBB-II. (a) Empirical MSEs of the bootstrap variance estimators of σ_n^2. The largest standard error is 0.058. (b) Ratio of the averaged variance estimates (over 3000 replications) to σ_n^2. The largest standard error is 0.097. (c) Coefficient of variation for 3000 independent estimates of σ_n^2. The largest standard error is 0.016 based on the jackknife estimate. (d) The empirical coverage probability of a two-sided 95% confidence interval of the mean. The largest standard error is 0.005.

expect that the aforementioned deficiency associated with the modified MBB carries over to the modified TBB when there are missing values.

Figure 2 compares the aforementioned 7 bootstrap schemes in the presence of 10 missing values. It appears that the DWB2 (DWB with tapering) and TBB-I (TBB with missing values ignored) perform the best of these, although the DWB2 is noticeably better (worse) than TBB-I when the block size is large (small), as seen from Figure 2(a) and (d). The imputation does affect the accuracy of variance estimation and empirical coverage somewhat. In addition, a comparison of MBBs with their tapered counterparts suggests that tapering is preferred. The modified MBB scheme (MBB-III) is inferior to the DWB2 uniformly in the examined block sizes and performs rather poorly for medium and large block sizes, as explained in the previous paragraph. We also compared the DWB2 (TBB-I) in Figure 2 and the DWB without missing values (TBB) in Figure 1 to demonstrate the impact of the missing observations (results not shown). It turns out that the presence of missing values had little affect on the performance of the DWB and TBB-I, which might be expected because the percentage of the missing data is only 5%.

We further compare the foregoing bootstrap schemes when the missing percentage is 25%. Note that none of the bootstrap methods (i.e., MBB, TBB, and DWB) has been theoretically justified for the situation where the missing percentage is large. Nevertheless, these methods are of practical use. In particular, we examined two missing patterns: (I) missing in blocks, with the 200 observations missing at the following 50 time points: $t = \{j, j+1, j+2, j+3, j+4\}$ for $j = 1, 21, 41, 61, 81, 101, 121, 141, 161, 181$, and (II) missing "irregularly," with the missing time points randomly generated from the uniform distribution over $\{1, \ldots, 200\}$. For our study, the missing time points are 1, 2, 7, 16, 24, 26, 27, 32, 33, 44, 59, 63, 68, 71, 88, 92, 94, 97, 98, 104, 115, 117, 118, 121, 134, 135, 136, 142, 143, 146, 147, 148, 149, 150, 152, 157, 161, 167, 169, 173, 174, 178, 183, 185, 186, 188, 189, 194, 195, and 198. Figures 3 and 4 correspond to (I) and (II), respectively. From these two figures, we again see that the DWB2 and TBB-I outperforms all of the other bootstrap schemes in all aspects. Because the percentage of missing is 25%, the bootstrap schemes with

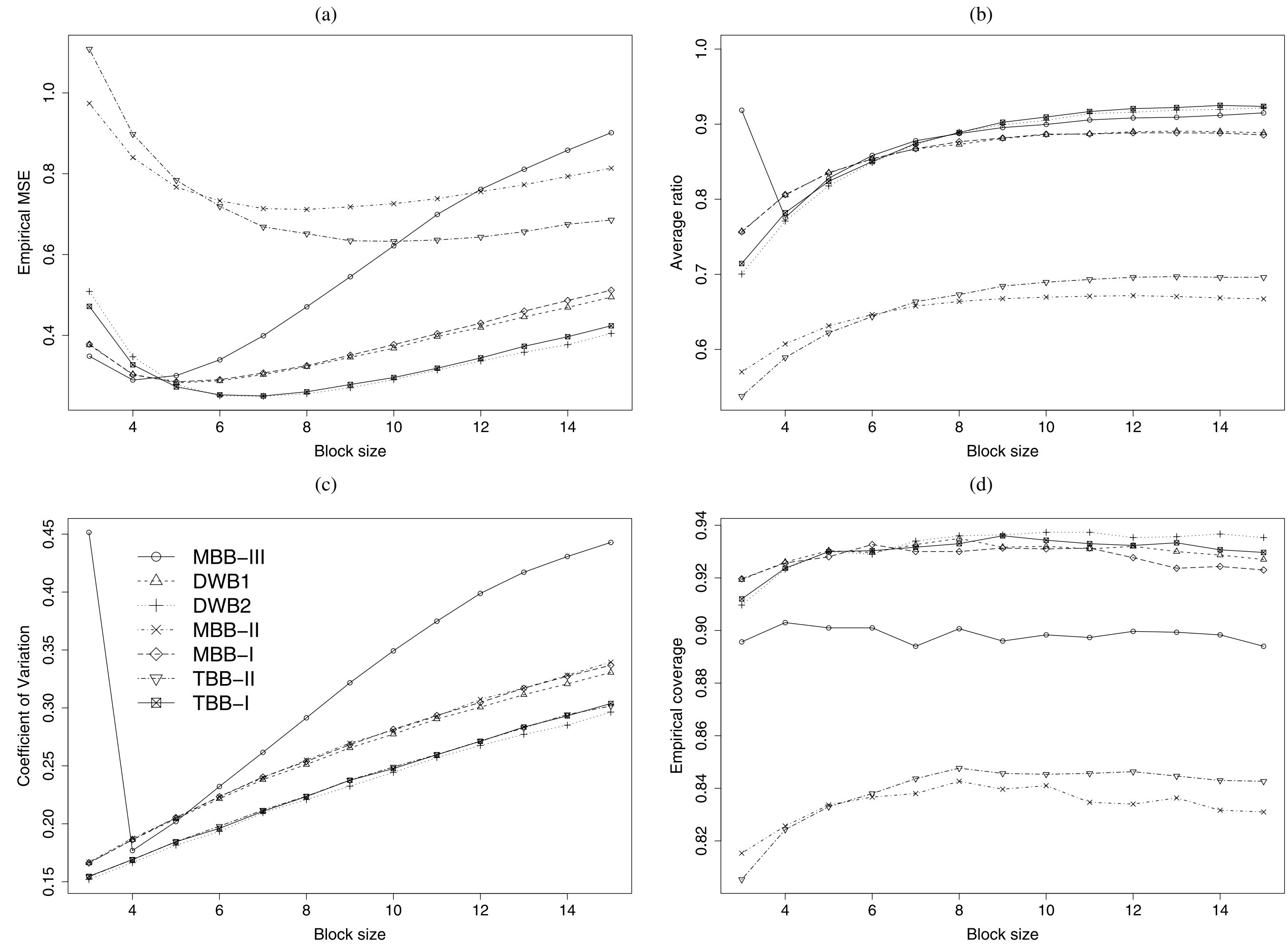

Figure 3. Time series generated from the model (9) with 50 missing values following the missing pattern (I) (i.e., missing in blocks); sample size, $n = 150$. The comparison is among MBB-I, MBB-II, MBB-III, DWB1, DWB2, TBB-I, and TBB-II. (a) Empirical MSEs of the bootstrap variance estimators of σ_n^2. The largest standard error is 0.026. (b) Ratio of the averaged variance estimates (over 3000 replications) to σ_n^2. The largest standard error is 0.111. (c) Coefficient of variation for 3000 independent estimates of σ_n^2. The largest standard error is 0.007 based on the jackknife estimate. (d) Empirical coverage probability of a two-sided 95% confidence interval of the mean. The largest standard error is 0.007.

imputed missing values deliver substantially worse results, and thus are not recommended. The modified MBB scheme is seen to perform very poorly for the missing pattern (II), with its erratic behavior presumably due to the particular missing pattern under consideration. The performance of the modified MBB is better for the missing pattern (I) but is still inferior to that of the DWB2 and TBB-I. Moreover, when the block size exceeds 15 for the missing pattern (I) [16 for the missing pattern (II)], the modified MBB scheme fails because there is no complete block to sample from. This could be a disadvantage if the optimal block size happens to be larger than the maximum size of a complete block with no missing values. Based on the foregoing simulation results, the TBB with ignored missing values and the DWB with tapering are the two schemes we would recommend in the case of lattice data with missing values. Note that implementing the TBB with ignored missing values is less straightforward than implementing the DWB, which in our experience is relatively easier to program. This implementational convenience for the DWB may be appealing to practitioners.

6.2 Time Series With Randomly Sampled Time Points

Here we investigate the finite-sample performance of the DWB under the framework of stochastic sampling design and compare it with that of the grid-based block bootstrap (Lahiri and Zhu 2006). Let $R_0 = (-1/2, 1/2]$, $n = 100$ and $\lambda_n = 18$ or 36. The density function for Z_1 is taken to be (a) truncated N(0, 1), that is, $\eta(x) = (2\pi)^{-1/2}\exp(-x^2/2)/\int_{-1/2}^{1/2}(2\pi)^{-1/2} \times \exp(-x^2/2)\,dx$ for $x \in (-1/2, 1/2]$ and 0 otherwise, or (b) truncated N(0, 1/4), that is, $\eta(x) = \mathbf{1}(|x| \le 1/2)2/\sqrt{2\pi}\exp(-2x^2)/\int_{-1/2}^{1/2} 2/\sqrt{2\pi}\exp(-2x^2)\,dx$. The truncated N(0, 1/4) distribution puts more mass around the origin than the truncated N(0, 1) distribution, which is close to a uniform distribution over R_0. Given the sampled time points $\{t_j\}_{j=1}^n$, we then generate the observations $\{X(t_j)\}_{j=1}^n$ from a one-dimensional mean-0 Gaussian process with exponential covariance function $\gamma(z) = \exp(-\rho|z|)$, $z \in \mathbb{R}$, where $\rho = 0.5$, 1, and 2. Table 1 shows the normalized MSEs in estimating $n\xi_n$ [i.e., $n\,\mathrm{var}(\bar{X}_n)$] for three bootstrap schemes: (a) the grid-based block bootstrap, (b) the DWB with $a(x) = w_{0.43}^{\mathrm{trap}} * w_{0.43}^{\mathrm{trap}}(x)/w_{0.43}^{\mathrm{trap}} * w_{0.43}^{\mathrm{trap}}(0)$ (i.e.,

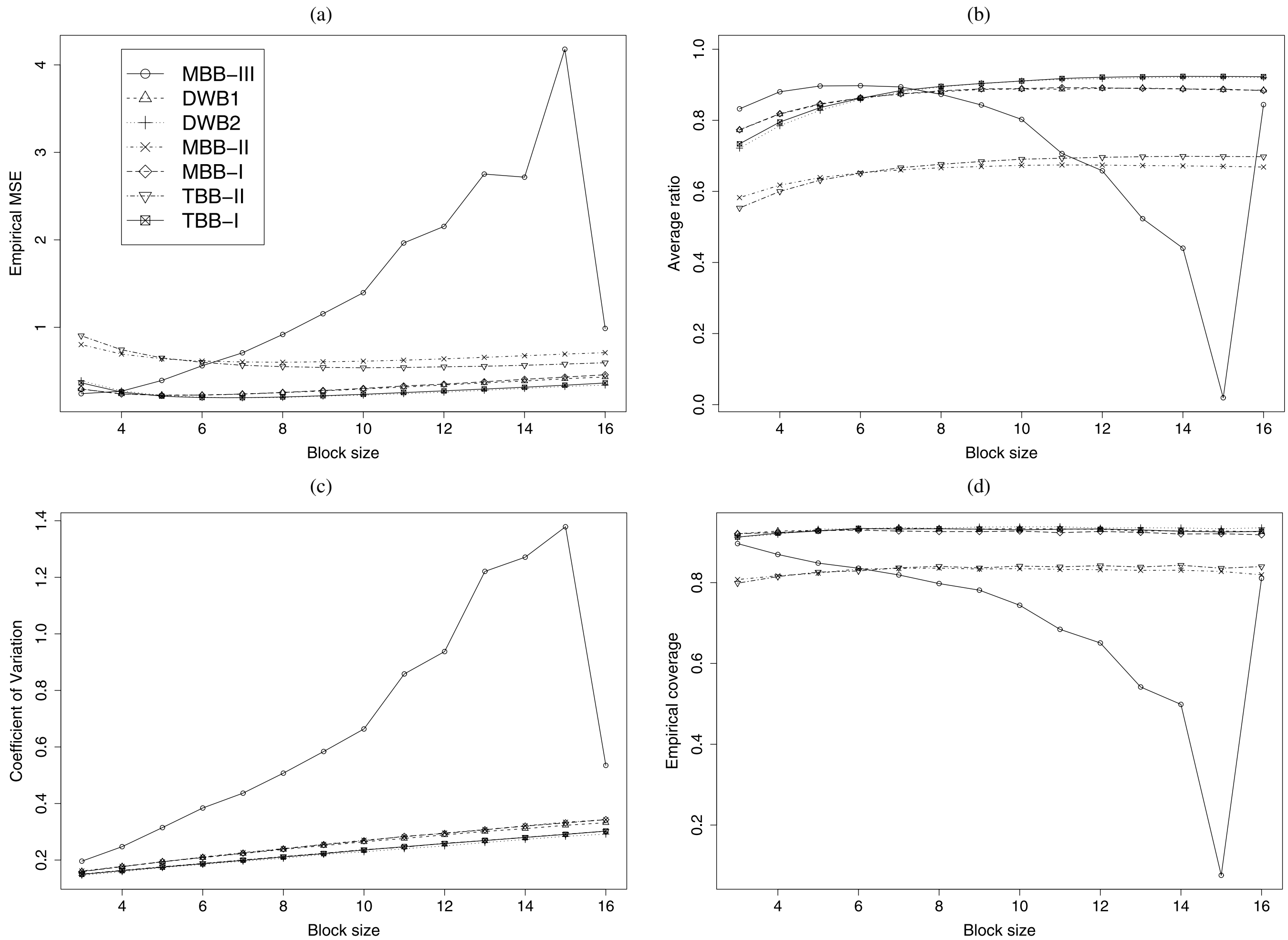

Figure 4. Time series generated from the model (9) with 50 missing values following the missing pattern (II); sample size, $n = 150$. The comparison is among MBB-I, MBB-II, MBB-III, DWB1, DWB2, TBB-I and TBB-II. (a) Empirical MSEs of the bootstrap variance estimators of σ_n^2. The largest standard error is 0.075. (b) Ratio of the averaged variance estimates (over 3000 replications) to σ_n^2. The largest standard error is 0.171. (c) Coefficient of variation for 3000 independent estimates of σ^2. The largest standard error is 0.028 based on the jackknife estimate. (d) Empirical coverage probability of a two-sided 95% confidence interval of the mean. The largest standard error is 0.005.

with tapering), and (c) the DWB with $a(x) = (1 - |x|)\mathbf{1}(|x| \le 1)$ (i.e., without tapering). Note that there is no straightforward extension of the TBB to the irregularly spaced case, so we include only the foregoing three schemes in the comparison. For any one of the bootstrap schemes, let $\hat{\xi}_n^{(j)}$ denote the bootstrap estimate of ξ_n based on 1000 bootstrap samples for the jth replicate, where $j = 1, \ldots, 1000$ because 1000 replications are used. Then the normalized MSE is calculated as $\sum_{j=1}^{1000}\{n\hat{\xi}_n^{(j)}/(n\xi_n) - 1\}^2/1000$. From Table 1, we can see that the optimal MSE for the DWB with tapering is slightly smaller than that for the grid-based block bootstrap and the DWB without tapering. This phenomenon can be seen for almost all combinations of $(\lambda_n, \rho, \eta(\cdot))$. Larger ρ corresponds to smaller MSE, which is expected because larger ρ implies weaker dependence. Moreover, the MSE decreases as λ_n increases. It appears that the advantage of the DWBs (with or without tapering) over the grid-based block bootstrap is noticeable for moderately large block sizes. Table 2 shows the empirical coverages in percentage in the same format as Table 1. Again, the DWB with tapering outperforms the other two bootstrap schemes in terms of optimal coverage. The grid-based block bootstrap performs poorly when l (block size in the grid-based block bootstrap) is large, although it has a slight edge over the DWB with tapering when l is small and suboptimal (e.g., $l = 1$). These findings are consistent with those for the regularly spaced time series with/without missing values presented in Section 6.1 and by Paparoditis and Politis (2001). We also tried the spherical covariance function $\gamma(z) = (1 - 3/2|z|/R + 1/2|z|^3/R^3)\mathbf{1}(|z| \le R)$ for $R = 4, 8$. As shown in the two tables in the online supplementary materials, the normalized MSEs and coverages exhibit the same pattern as described earlier. Qualitatively similar behaviors for the three bootstrap schemes when $n = 200$ are observed, but for space considerations we do not discuss these results here.

In the foregoing simulation studies, we do not consider the issue of bandwidth selection, which is very important in practical data analysis. There is a large literature on selecting the bandwidth l_n for regularly spaced time series (see Lahiri 2003a, chap. 7 for a review). The two major approaches—the nonpara-

Table 1. Normalized MSEs for the bootstrap variance estimators of $n\,\mathrm{var}(\bar{X}_n)$ using (a) the grid-based block bootstrap, (b) the DWB with $a(x) = w_{0.43}^{\mathrm{trap}} * w_{0.43}^{\mathrm{trap}}(x)/w_{0.43}^{\mathrm{trap}} * w_{0.43}^{\mathrm{trap}}(0)$, and (c) the DWB with $a(x) = (1 - |x|)\mathbf{1}(|x| \le 1)$. The data are mean-0 Gaussian with an exponential covariance function. The box for each row indicates the smallest normalized MSE among $l = 1, \ldots, 10$. Part (A) corresponds to the truncated N(0, 1) density function for the sampling design, whereas part (B) is for the truncated N(0, 1/4) density function. The largest standard error is 0.020 for part (A) and 0.026 for part (B)

			l									
λ_n	ρ		1	2	3	4	5	6	7	8	9	10
(A)												
18	0.5	(a)	0.64	0.51	**0.48**	0.48	0.50	0.54	0.55	0.60	0.66	0.66
		(b)	0.69	0.56	0.49	0.46	**0.44**	0.45	0.46	0.47	0.48	0.50
		(c)	0.65	0.52	0.47	**0.46**	0.47	0.49	0.50	0.52	0.54	0.55
	1	(a)	0.43	**0.34**	0.34	0.37	0.41	0.48	0.48	0.54	0.62	0.62
		(b)	0.49	0.35	**0.31**	0.32	0.33	0.36	0.38	0.40	0.43	0.46
		(c)	0.44	0.34	**0.33**	0.35	0.38	0.42	0.44	0.47	0.49	0.51
	2	(a)	0.25	**0.23**	0.28	0.32	0.37	0.45	0.44	0.48	0.57	0.56
		(b)	0.27	**0.21**	0.23	0.26	0.30	0.33	0.36	0.39	0.41	0.44
		(c)	0.25	**0.23**	0.26	0.31	0.34	0.37	0.40	0.42	0.44	0.46
36	0.5	(a)	0.58	0.44	0.37	**0.34**	0.34	0.35	0.37	0.38	0.41	0.42
		(b)	0.63	0.49	0.40	0.36	0.33	**0.32**	0.32	0.32	0.33	0.34
		(c)	0.58	0.44	0.37	0.34	**0.33**	0.34	0.35	0.37	0.38	0.40
	1	(a)	0.35	0.25	**0.22**	0.23	0.25	0.28	0.30	0.32	0.36	0.37
		(b)	0.40	0.27	0.22	**0.21**	0.21	0.22	0.24	0.26	0.28	0.29
		(c)	0.36	0.25	**0.22**	0.23	0.24	0.26	0.29	0.31	0.33	0.35
	2	(a)	0.18	**0.15**	0.17	0.19	0.22	0.25	0.27	0.29	0.34	0.34
		(b)	0.20	**0.14**	0.15	0.16	0.18	0.20	0.22	0.24	0.26	0.28
		(c)	0.18	**0.15**	0.16	0.19	0.21	0.23	0.26	0.28	0.31	0.33
(B)												
18	0.5	(a)	0.70	0.59	**0.56**	0.57	0.59	0.63	0.66	0.72	0.80	0.80
		(b)	0.74	0.63	0.56	0.54	**0.53**	0.53	0.54	0.56	0.58	0.59
		(c)	0.70	0.59	0.56	**0.55**	0.57	0.59	0.61	0.63	0.66	0.68
	1	(a)	0.48	**0.40**	0.40	0.43	0.47	0.53	0.56	0.64	0.74	0.74
		(b)	0.53	0.40	**0.37**	0.37	0.39	0.41	0.43	0.46	0.49	0.51
		(c)	0.49	0.40	**0.39**	0.42	0.45	0.48	0.51	0.54	0.57	0.60
	2	(a)	**0.28**	0.28	0.32	0.36	0.40	0.48	0.50	0.58	0.70	0.71
		(b)	0.31	**0.26**	0.28	0.31	0.34	0.37	0.40	0.42	0.45	0.48
		(c)	0.29	**0.28**	0.32	0.36	0.39	0.42	0.45	0.48	0.52	0.54
36	0.5	(a)	0.61	0.49	0.43	**0.41**	0.41	0.42	0.43	0.44	0.47	0.48
		(b)	0.66	0.53	0.45	0.41	0.39	**0.38**	0.39	0.39	0.40	0.41
		(c)	0.62	0.48	0.42	**0.40**	0.40	0.41	0.42	0.43	0.45	0.46
	1	(a)	0.38	0.29	**0.27**	0.29	0.31	0.33	0.35	0.37	0.40	0.41
		(b)	0.43	0.30	**0.26**	0.26	0.27	0.29	0.31	0.32	0.34	0.35
		(c)	0.38	0.29	**0.27**	0.28	0.31	0.33	0.34	0.36	0.38	0.39
	2	(a)	0.21	**0.19**	0.22	0.25	0.28	0.32	0.34	0.36	0.40	0.40
		(b)	0.23	**0.18**	0.19	0.22	0.25	0.28	0.31	0.34	0.36	0.38
		(c)	0.21	**0.19**	0.22	0.26	0.30	0.32	0.35	0.37	0.38	0.40

metric plug-in method (e.g., Bülmann and Künsch 1999; Paparoditis and Politis 2001, 2002; Politis and White 2004; Lahiri, Furukawa, and Lee 2007) and the subsampling method (Hall, Horowitz, and Jing 1995)—can be extended to the DWB for time series on a lattice; see Section 6.3 for the use of the plug-in method. For a nonlattice temporal series, it seems that no bandwidth selection method is available for either the grid-based block bootstrap or the DWB.

Table 2. Empirical coverages (in percent) for the bootstrap-based confidence intervals of μ using (a) the grid-based block bootstrap, (b) the DWB with $a(x) = w^{\mathrm{trap}}_{0.43} * w^{\mathrm{trap}}_{0.43}(x)/w^{\mathrm{trap}}_{0.43} * w^{\mathrm{trap}}_{0.43}(0)$, and (c) the DWB with $a(x) = (1 - |x|)\mathbf{1}(|x| \le 1)$. The data are mean-0 Gaussian with an exponential covariance function. The box for each row indicates the best coverage among $l = 1, \ldots, 10$ (i.e., closest to the nominal level 95%). The largest standard error is 1.6%. Part (A) corresponds to the truncated N(0, 1) density function for the sampling design, whereas part (B) is for the truncated N(0, 1/4) density function

			l									
λ_n	ρ		1	2	3	4	5	6	7	8	9	10
(A)												
18	0.5	(a)	61	68	**70**	69	68	64	64	59	53	54
		(b)	58	66	69	71	71	**72**	71	70	69	68
		(c)	62	68	69	**70**	70	69	68	67	65	63
	1	(a)	73	**79**	79	78	76	70	71	66	58	57
		(b)	70	78	81	**82**	81	81	80	78	76	76
		(c)	73	79	**81**	80	78	77	75	73	70	69
	2	(a)	82	**84**	83	79	77	72	72	68	60	60
		(b)	82	**86**	86	85	84	83	81	80	78	76
		(c)	83	**85**	84	83	81	79	77	76	74	73
36	0.5	(a)	62	72	76	**78**	78	78	76	76	74	73
		(b)	59	69	74	77	79	79	**80**	80	80	79
		(c)	63	71	76	78	**79**	79	78	78	77	77
	1	(a)	76	83	84	**85**	84	83	81	81	77	77
		(b)	74	81	84	**86**	86	86	86	86	85	85
		(c)	76	83	**85**	85	85	84	84	83	82	82
	2	(a)	85	**88**	88	87	85	85	83	82	79	79
		(b)	85	**89**	89	89	89	88	88	87	87	86
		(c)	86	88	**89**	88	87	87	85	85	84	83
(B)												
18	0.5	(a)	53	60	**62**	61	60	56	55	49	41	41
		(b)	51	60	64	66	**67**	67	66	64	62	60
		(c)	54	62	**65**	65	64	61	59	56	54	52
	1	(a)	70	**75**	74	73	71	66	65	59	48	49
		(b)	68	76	**78**	78	78	77	74	72	70	68
		(c)	71	76	**77**	76	73	70	68	66	64	62
	2	(a)	80	**82**	80	78	75	70	70	64	52	53
		(b)	80	84	**86**	84	83	80	79	76	73	70
		(c)	82	**84**	83	80	78	76	72	70	68	66
36	0.5	(a)	61	68	70	**72**	72	71	70	69	68	67
		(b)	58	67	71	73	74	74	74	**75**	74	74
		(c)	61	68	72	**73**	73	73	73	71	70	69
	1	(a)	76	80	**81**	81	80	78	78	76	74	74
		(b)	74	80	83	**84**	84	84	83	82	81	80
		(c)	76	82	**83**	83	82	81	80	79	77	77
	2	(a)	83	**86**	86	84	83	81	79	78	75	75
		(b)	83	86	**87**	87	87	86	85	84	83	81
		(c)	85	**87**	86	86	84	82	82	80	79	77

Several issues merit further investigation. For the DWB, we need to choose the bandwidth l, the covariance kernel $a(\cdot)$, and the joint distribution of W_t. The joint distribution of $\{W_t\}_{t=1}^n$ is taken to be multivariate normal in our simulations and empirical data analysis. It would be interesting to explore how the results are sensitive to the specification of the joint distribution; see Example 4.1. In practice, if a data-driven bandwidth is used, then a natural question is how much improvement can the

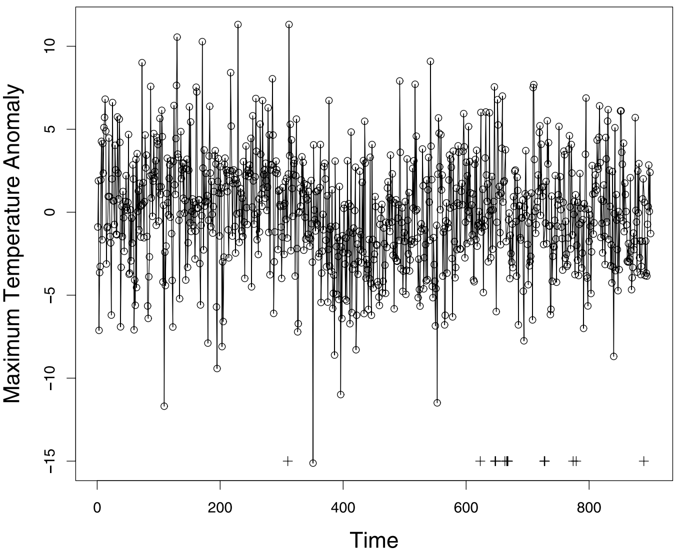

Figure 5. The monthly maximum temperature anomaly (F) recorded at Salisbury, North Carolina during 1931–2005. The symbol "+" indicates the missing values.

DWB offer compared with the MBB when coupled with different bandwidth selection algorithms. In addition, more time series models and smooth functionals (see Sec. 6.3 for an example) need to be examined through simulations to get a broad picture of the finite-sample performance of the DWB method. Given our space limitations here, we prefer to address these issues in a separate numerical work.

6.3 Empirical Illustration

To illustrate the usefulness of the DWB, we analyze the monthly maximum temperature anomaly at Salisbury, North Carolina for the period 1931–2005. The data set can be obtained at *http://lwf.ncdc.noaa.gov/oa/climate/research/ushcn/ushcn.html*, which stores high-quality moderate sized data set of monthly averaged maximum, minimum, and mean temperature anomalies for more than 1000 stations in the United States. Figure 5 shows the times series plot with the symbol "+" indicating the missing values, which occur at the following months: 10/1956, 11/1982, 11/1984, 12/1984, 3/1986, 6/1986, 7/1986, 8/1986, 7/1991, 8/1991, 6/1995, 11/1995, and 1/2005. Thus the sample size is $n = 887$ with 13 missing values.

At the initial stage of model building, it is an important step to compute the empirical autocorrelations at a number of lags and assess if they are significantly different from 0. Toward this end, we construct 95% confidence intervals for $\rho_k = \gamma_k/\gamma_0$, $k = 1, \ldots, 5$. Note that the asymptotic variance for the sample estimator $\hat{\rho}_k$ involves the integral of the fourth-order cumulants, and a consistent estimation of its asymptotic variance is very involved even for regularly spaced time series. One way of bypassing direct estimation of asymptotic variance is to use the subsampling and block-based bootstrap methods, as advocated by Romano and Thombs (1996). Here we apply the DWB method to this data set; the implementation is quite straightforward. Following the discussion in example 4.7 of Lahiri (2003a), we can write $\rho_k = H(\boldsymbol{\mu})$, where $\boldsymbol{\mu} = \mathbb{E}(\mathbf{Y}_1)$, $\mathbf{Y}_i = (Y_{1i}, Y_{2i}, Y_{3i})' = (X_i, X_i^2, X_iX_{i+k})'$, and $H(x, y, z) = \{(z - x^2)/(y - x^2)\}\mathbf{1}(y > x^2)$ for $(x, y, z)' \in \mathbb{R}^3$. Let $S_k = \{j = 1, \ldots, n : X_j$ and X_{j+k} are both nonmissing$\}$, $k \in \mathbb{N}$, $\bar{Y}_{jn} = |S_k|^{-1} \sum_{i\in S_k} Y_{ji}, j = 1, 2, 3$, and $\bar{\mathbf{Y}}_n = (\bar{Y}_{1n}, \bar{Y}_{2n}, \bar{Y}_{3n})'$. We consider the following sample estimate of ρ_k:

$$\hat{\rho}_k = \frac{|S_k|^{-1} \sum_{j\in S_k} X_jX_{j+k} - (|S_k|^{-1} \sum_{j\in S_k} X_j)^2}{|S_k|^{-1} \sum_{j\in S_k} X_j^2 - (|S_k|^{-1} \sum_{j\in S_k} X_j)^2} = H(\bar{Y}_{1n}, \bar{Y}_{2n}, \bar{Y}_{3n}).$$

To implement the DWB, we use $a(x) = w_{0.43}^{\text{trap}} * w_{0.43}^{\text{trap}}(x)/w_{0.43}^{\text{trap}} * w_{0.43}^{\text{trap}}(0)$ and select the bandwidth l using the plug-in method suggested by Paparoditis and Politis (2001) and Politis and White (2004), which has been found to work well for block-based methods. We estimate B_0 and D_0 in the expression of l_n^{opt} using the flat-top window $\lambda(t) = \mathbf{1}_{[0,1/2]}(|t|) + 2(1 - |t|)\mathbf{1}_{(1/2,1]}(|t|)$, $t \in [-1, 1]$ (see Politis and Romano 1995). Specifically, we estimate $\sum_{k=-\infty}^{\infty} k^2 r_k$ by $\hat{\mathbf{\nabla}}' \sum_{k=-2M}^{2M} \lambda\{k/(2M)\}\hat{\boldsymbol{\gamma}}_{Yk} k^2 \hat{\mathbf{\nabla}}$, where $\hat{\mathbf{\nabla}} = \partial H(\bar{\mathbf{Y}}_n)/\partial\boldsymbol{\mu}$, $\hat{\boldsymbol{\gamma}}_{Yk} = |S_{|k|}|^{-1} \times \sum_{j\in S_{|k|}} (\mathbf{Y}_j - \bar{\mathbf{Y}}_n)(\mathbf{Y}_{j+|k|} - \bar{\mathbf{Y}}_n)'$, and M is a positive integer. Similarly, the quantity $\tau_\infty^2 = \sum_{k=-\infty}^{\infty} r_k$ is estimated by $\hat{\mathbf{\nabla}}' \sum_{k=-2M}^{2M} \lambda\{k/(2M)\}\hat{\boldsymbol{\gamma}}_{Yk}\hat{\mathbf{\nabla}}$. The data-driven optimal bandwidth l_n^{opt} is then obtained by plugging these estimates into the expressions for B_0 and D_0. For the choice of M, Politis and Romano (1995) suggested taking M to be the smallest integer after which the correlogram appears negligible. In our setting, $\mathbf{Y}_t$ is a three-dimensional time series, and the autocorrelations for each component are reasonably small when the lag is beyond 1. Table 3 shows the 95% confidence interval of ρ_k for a range of M.

It appears that the confidence intervals do not vary much for $M \in \{1, \ldots, 6\}$. It seems safe to conclude that (a) there are significant nonzero autocorrelations at the first two lags, (b) the significance is marginal at lags 3 and 4, and (c) no sufficient evidence is available to support $\rho_5 \neq 0$. Note that 3000 bootstrap samples are used in the calculation.

Table 3. The 95% confidence intervals for $\{\rho_k\}_{k=1}^5$ with $M = 1, \ldots, 6$

$M\backslash(k\|\hat{\rho}_k)$	1‖0.272	2‖0.080	3‖0.058	4‖0.050	5‖0.016
1	[0.219, 0.333]	[0.012, 0.153]	[−0.006, 0.126]	[−0.010, 0.115]	[−0.050, 0.083]
2	[0.217, 0.332]	[0.012, 0.149]	[−0.007, 0.125]	[−0.011, 0.117]	[−0.045, 0.079]
3	[0.220, 0.329]	[0.015, 0.148]	[−0.008, 0.127]	[−0.010, 0.113]	[−0.046, 0.084]
4	[0.221, 0.327]	[0.023, 0.149]	[−0.004, 0.128]	[−0.006, 0.111]	[−0.045, 0.079]
5	[0.221, 0.331]	[0.022, 0.149]	[−0.004, 0.128]	[−0.006, 0.115]	[−0.044, 0.081]
6	[0.223, 0.328]	[0.022, 0.149]	[−0.005, 0.127]	[−0.005, 0.110]	[−0.049, 0.081]

7. CONCLUSION

We have proposed a new resampling method for time series, the DWB, that is generally applicable to variance estimation and sampling distribution approximation for the smooth function model. Compared with existing block-based bootstrap methods, the DWB has a number of appealing features:

1. For variance estimation of regularly spaced time series, the DWB is asymptotically equivalent to the TBB (with proper choice of the covariance kernel), which outperforms all other block-based methods in terms of the bias and MSE. As shown in our simulation results, this advantage for the DWB carries over to irregularly spaced time series, for which the TBB has no straightforward extension.
2. Computationally, it is very convenient to implement the DWB if the joint distribution of W_t is chosen to be multivariate normal or its simple variants. In addition, the bandwidth in the DWB does not have to be an integer, so the optimal theoretical MSE in the variance estimation is achievable in practice.
3. A common undesirable feature of block-based bootstrap methods is that if the sample size is not a multiple of the block size, then one must either take a shorter bootstrap sample or use a fraction of the last resampled block. This could lead to some inaccuracy when the block size is large. In contrast, for the DWB, the size of the bootstrap sample is always same as the original sample size.
4. The DWB can be easily extended to the spatial setting, where irregular spaced data seems to be the rule rather than the exception. This is currently under investigation.

A major advantage of the DWB over the block-based bootstrap methods is its convenience to implement for irregularly spaced data. Having said that, we also should note that for regularly spaced time series, the DWB is not as widely applicable as the MBB, and the DWB lacks the higher order accuracy property of the MBB. It is our view that the DWB is a complement to, but not a competitor of, existing block-based bootstrap methods. In addition, we are well aware of other bootstrap methods in time series, such as parametric bootstrap (i.e., bootstrap assuming parametric time series models), sieve bootstrap (Bühlmann 1997), and frequency domain bootstrap (Franke and Härdle 1992). But they are inconvenient to use in the presence of unequally spaced time points. Because the DWB has its own advantages not shared by all of the existing bootstrap methods developed for regularly spaced time series, it can be recommended to the practitioner as an additional tool for the inference of irregularly spaced dependent data.

APPENDIX: PROOFS

Proof of Theorem 3.1

Let $\Phi(\mathbf{x};\boldsymbol{\Sigma}_\infty)$ be the distribution function of $\mathrm{N}(\mathbf{0},\boldsymbol{\Sigma}_\infty)$ on $\mathbb{R}^p$. We first show that

$$\sup_{\mathbf{x}\in\mathbb{R}^p}\left|P\{\sqrt{n}(\bar{\mathbf{X}}_n-\boldsymbol{\mu})\le\mathbf{x}\}-P^*\{\sqrt{n}(\bar{\mathbf{X}}^*_{n,DWB}-\bar{\mathbf{X}}_n)\le\mathbf{x}\}\right| = o_p(1). \quad \text{(A.1)}$$

Because $\sqrt{n}(\bar{\mathbf{X}}_n-\boldsymbol{\mu})\to_D \mathrm{N}(\mathbf{0},\boldsymbol{\Sigma}_\infty)$ under Assumption 3.1, it follows from a multivariate version of Polyā's theorem (Bhattacharya and Rao 1986) that $\sup_{\mathbf{x}\in\mathbb{R}^p}|P\{\sqrt{n}(\bar{\mathbf{X}}_n-\boldsymbol{\mu})\le\mathbf{x}\}-\Phi(\mathbf{x};\boldsymbol{\Sigma}_\infty)|=o(1)$. Then (A.1) follows if we can show that

$$\sup_{\mathbf{x}\in\mathbb{R}^p}\left|P^*\{\sqrt{n}(\bar{\mathbf{X}}^*_{n,DWB}-\bar{\mathbf{X}}_n)\le\mathbf{x}\}-\Phi(\mathbf{x};\boldsymbol{\Sigma}_\infty)\right|=o_p(1).$$

Write $\sqrt{n}(\bar{\mathbf{X}}^*_{n,DWB}-\bar{\mathbf{X}}_n)=n^{-1/2}\sum_{t=1}^n W_t(\mathbf{X}_t-\boldsymbol{\mu})+n^{-1/2}\times\sum_{t=1}^n W_t(\boldsymbol{\mu}-\bar{\mathbf{X}}_n)=\mathbf{M}_n+\mathbf{R}_n$. It is easily seen that $\mathbb{E}^*(\|\mathbf{R}_n\|^2)=(\bar{\mathbf{X}}_n-\boldsymbol{\mu})'(\bar{\mathbf{X}}_n-\boldsymbol{\mu})O(l_n)=o_p(1)$. Let $\mathbf{Y}_\infty\sim\mathrm{N}(\mathbf{0},\boldsymbol{\Sigma}_\infty)$. By lemma 4.1 of Lahiri (2003a) and the Cramer–Wold device, it suffices to show that for any $\mathbf{h}\in\mathbb{R}^p$,

$$\mathbf{h}'\mathbf{M}_n=n^{-1/2}\sum_{t=1}^n W_t\mathbf{h}'(\mathbf{X}_t-\boldsymbol{\mu})\to_D \mathbf{h}'\mathbf{Y}_\infty, \quad \text{(A.2)}$$

in probability. Letting $Y_t=\mathbf{h}'(\mathbf{X}_t-\boldsymbol{\mu})$, $\mathrm{var}^*(\mathbf{h}'\mathbf{M}_n)=n^{-1}\times\sum_{t,t'=1}^n Y_tY_{t'}a(|t-t'|/l)$. According to Proposition 2.1 and the references therein, we have $\mathbb{E}\{\mathrm{var}^*(\mathbf{h}'\mathbf{M}_n)\}=\mathbf{h}'\boldsymbol{\Sigma}_\infty\mathbf{h}+o(1)$ and $\mathrm{var}\{\mathrm{var}^*(\mathbf{h}'\mathbf{M}_n)\}=O(l/n)$. So we get $\mathrm{var}^*(\mathbf{h}'\mathbf{M}_n)\to_p\mathbf{h}'\boldsymbol{\Sigma}_\infty\mathbf{h}$ as $n\to\infty$.

In view of the l-dependence of W_t, we adopt the large-block–small-block argument to prove the central limit theorem (A.2). Define a sequence of numbers $\{L_n\}$ that satisfy $L_n\to\infty, l_n=o(L_n)$, and $k_n=\lfloor n/(L_n+l_n)\rfloor\to\infty$. Define the blocks $\mathcal{L}_r=\{j\in\mathbb{N}:(r-1)(L_n+l_n)+1\le j\le r(l_n+L_n)-l_n\}$, $1\le r\le k_n$, $\mathcal{S}_r=\{j\in\mathbb{N}:r(L_n+l_n)-l_n+1\le j\le r(l_n+L_n)\}$, $1\le r\le k_n-1$, and $\mathcal{S}_{k_n}=\{j\in\mathbb{N}:k_n(L_n+l_n)-l_n+1\le j\le n\}$. Let $U_r=\sum_{j\in\mathcal{L}_r}W_jY_j$ and $V_r=\sum_{j\in\mathcal{S}_r}W_jY_j$, $r=1,\dots,k_n$. Conditional on $\mathcal{X}_n$, $U_1,\dots,U_{k_n}$ are independent random variables, as are $V_1,\dots,V_{k_n-1}$. Note that for $r=1,\dots,k_n-1$, $\mathbb{E}^*(V_r^2)=\sum_{j,j'\in\mathcal{S}_r}Y_jY_{j'}a(|j-j'|/l)\ge 0$, which implies that $\mathbb{E}\{\mathbb{E}^*(V_r^2)\}=\sum_{j,j'\in\mathcal{S}_r}\mathbf{h}'\,\mathrm{cov}(\mathbf{X}_j,\mathbf{X}_{j'})\mathbf{h}\cdot a(|j-j'|/l)\le Cl_n$. A similar argument yields $\mathbb{E}\{\mathbb{E}^*(V_{k_n}^2)\}=O(L_n)$. Thus $\mathbb{E}[\mathbb{E}^*(\sum_{r=1}^{k_n}V_r)^2\}=\sum_{r=1}^{k_n}\mathbb{E}\{\mathbb{E}^*(V_r^2)\}=O(k_nl_n+L_n)=o(n)$. Therefore, (A.2) follows if we can show that $n^{-1/2}\sum_{r=1}^{k_n}U_r\to_D\mathbf{h}'\mathbf{Y}_\infty$ in probability. For any $\epsilon>0$, let $\hat{\Delta}_n(\epsilon)=n^{-1}\sum_{r=1}^{k_n}\mathbb{E}^*\{U_r^2\times\mathbf{1}(|U_r|>\sqrt{n}\epsilon)\}$. Write $\|X\|_p^*=(\mathbb{E}^*|X|^p)^{1/p}$ for $p\ge1$. By the triangle inequality and Rosenthal inequality (conditional on $\mathcal{X}_n$), we obtain

$$\begin{aligned}\|U_1\|_{2+\delta}^* &\le \sum_{g=1}^{2l}\left\|\sum_{j=1}^{\lfloor (L_n-g)/(2l)\rfloor}W_{g+2(j-1)l}Y_{g+2(j-1)l}\right\|_{2+\delta}^*\\ &\le C\sum_{g=1}^{2l}\left\{\left\|\sum_{j=1}^{\lfloor (L_n-g)/(2l)\rfloor}W^2_{g+2(j-1)l}Y^2_{g+2(j-1)l}\right\|^*_{1+\delta/2}\right\}^{1/2}\\ &\le C\sum_{g=1}^{2l}\left\{\sum_{j=1}^{\lfloor (L_n-g)/(2l)\rfloor}\|W^2_{g+2(j-1)l}\|^*_{1+\delta/2}Y^2_{g+2(j-1)l}\right\}^{1/2}\\ &\le C\sum_{g=1}^{2l}\left\{\sum_{j=1}^{\lfloor (L_n-g)/(2l)\rfloor}Y^2_{g+2(j-1)l}\right\}^{1/2}\\ &\le C\sqrt{l}\left\{\sum_{g=1}^{2l}\sum_{j=1}^{\lfloor (L_n-g)/(2l)\rfloor}Y^2_{g+2(j-1)l}\right\}^{1/2}\\ &= C\sqrt{l}\left(\sum_{t\in\mathcal{L}_1}Y_t^2\right)^{1/2}, \end{aligned} \quad \text{(A.3)}$$

where the last inequality is due to the Cauchy–Schwarz inequality. By the same argument, $\|U_r\|^*_{2+\delta}\le C\sqrt{l}(\sum_{t\in\mathcal{L}_r}Y_t^2)^{1/2}$, $r=2,\dots,k_n$.

Applying Hölder's inequality,

$$\sum_{r=1}^{k_n}\mathbb{E}^*|U_r|^{2+\delta} \le Cl^{(1+\delta/2)}\sum_{r=1}^{k_n}\Big(\sum_{t\in\mathcal{L}_r} Y_t^2\Big)^{(1+\delta/2)}$$

$$\le Cl^{(1+\delta/2)}\sum_{r=1}^{k_n} L_n^{\delta/2}\sum_{t\in\mathcal{L}_r}|Y_t|^{2+\delta}$$

$$= O_p\big(l^{(1+\delta/2)}L_n^{\delta/2}n\big).$$

Therefore, if we let $L_n = n^{1/2}/l_n^{1/\delta}$, then, because $l_n = o\{n^{\delta/(2+2\delta)}\}$, we have $\hat{\Delta}_n(\epsilon) = o_p(1)$. Thus the assertion (A.1) holds in view of the argument in the proof of theorem 3.2 of Lahiri (2003a). Finally, our conclusion follows from the argument in the proof of theorem 4.1 of Lahiri (2003a). We omit the details here.

Proof of Theorem 5.1

We treat only case (a) here, because the proof for case (b) largely repeats that for case (a). Write

$$\hat{\xi}_n = n^{-2}\sum_{j,j'=1}^{n}\{X(t_j)-\bar{X}_n\}\{X(t_{j'})-\bar{X}_n\}a\{(t_j-t_{j'})/l_n\}$$

$$= n^{-2}\sum_{j,j'=1}^{n}\{X(t_j)-\mu\}\{X(t_{j'})-\mu\}a\{(t_j-t_{j'})/l_n\}$$

$$+ \sum_{j,j'=1}^{n}\{X(t_j)-\mu\}a\{(t_j-t_{j'})/l_n\}2n^{-2}(\mu-\bar{X}_n)$$

$$+ n^{-2}\sum_{j,j'=1}^{n}a\{(t_j-t_{j'})/l_n\}(\mu-\bar{X}_n)^2$$

$$= J_{1n}+J_{2n}+J_{3n}.$$

The conclusion follows if we can show

$$\mathbb{E}(nJ_{1n}) \to \gamma(0)+\kappa\iota\int_{\mathbb{R}}\gamma(s)\,ds, \qquad (A.4)$$

$$\mathrm{var}(nJ_{1n}) \to 0, \qquad (A.5)$$

$$nJ_{3n} = o_p(1), \qquad (A.6)$$

$$nJ_{2n} = o_p(1). \qquad (A.7)$$

To show (A.4), we note that

$$\mathbb{E}(nJ_{1n}) = \mathbb{E}_Z\{\mathbb{E}_{X|Z}(nJ_{1n})\}$$

$$= \gamma(0)+n^{-1}\sum_{j\ne j'=1}^{n}\mathbb{E}_Z[\gamma(t_j-t_{j'})a\{(t_j-t_{j'})/l_n\}]$$

$$= \gamma(0) + (n-1)\int_{R_0^2}\gamma\{\lambda_n(x-y)\} \times a\{\lambda_n/l_n(x-y)\}\eta(x)\eta(y)\,dx\,dy.$$

Let $Q_n = \int_{R_0^2}\gamma\{\lambda_n(x-y)\}a\{\lambda_n/l_n(x-y)\}\eta(x)\eta(y)\,dx\,dy$, $R_D = \{x-y: x\in R_0, y\in R_0\}$ and for any $z\in R_D$, $R(z) = (R_0+z)\cap R_0$. Write $Q_n = \int_{R_D}\int_{x\in R(z)}\eta(x)\eta(x-z)\,dx\,\gamma(\lambda_n z)a(\lambda_n z/l_n)\,dz = \lambda_n^{-1}\times\int_{\lambda_n R_D}\int_{x\in R(z/\lambda_n)}\eta(x)\eta(x-z/\lambda_n)\,dx\,\gamma(z)a(z/l_n)\,dz$. We divide the integral $\int_{\lambda_n R_D}$ into two parts $\int_{\lambda_n R_D\cap\{|z|\le M\}}$ and $\int_{\lambda_n R_D\cap\{|z|>M\}}$ for some $M>0$. When $z\in\lambda_n R_D\cap\{|z|\le M\}$, we have $a(z/l_n) = 1+o(1)$ uniformly over $|z|\le M$, which implies that

$$\int_{\lambda_n R_D\cap\{|z|\le M\}}\int_{x\in R(z/\lambda_n)}\eta(x)\eta(x-z/\lambda_n)\,dx\,\gamma(z)a(z/l_n)\,dz \to \iota\int_{|z|\le M}\gamma(z)\,dz.$$

When $z\in\lambda_n R_D\cap\{|z|>M\}$, we take advantage of the boundedness of $a(\cdot)$ and $\eta(\cdot)$ (on R_0) and get

$$\left|\int_{\lambda_n R_D\cap\{|z|>M\}}\int_{x\in R(z/\lambda_n)}\eta(x)\eta(x-z/\lambda_n)\,dx\,\gamma(z)a(z/l_n)\,dz\right| \le C\int_{|z|\ge M}|\gamma(z)|\,dz,$$

which can be made arbitrarily small by choosing a large M in view of (7). This implies that $\lambda_n Q_n \to \iota\int_{\mathbb{R}}\gamma(z)\,dz$ and, subsequently, that $\mathbb{E}(nJ_{1n}) \to \gamma(0)+\kappa\iota\int_{\mathbb{R}}\gamma(z)\,dz$ for case (a) and $\mathbb{E}(\lambda_n J_{1n}) \to \iota\int_{\mathbb{R}}\gamma(z)\,dz$ for case (b). Following the same argument, we can show that

$$n\xi_n \to \gamma(0)+\kappa\iota\int_{\mathbb{R}}\gamma(s)\,ds \quad\text{for case (a) and}$$
$$\lambda_n\xi_n \to \iota\int_{\mathbb{R}}\gamma(s)\,ds \quad\text{for case (b),} \qquad (A.8)$$

a fact that we make use of later.

To show (A.5), we note that $\mathrm{var}(nJ_{1n}) = \mathbb{E}_Z\{\mathrm{var}_{X|Z}(nJ_{1n})\} + \mathrm{var}_Z\{\mathbb{E}_{X|Z}(nJ_{1n})\}$ and show that each term approaches 0 as $n\to\infty$. Write $\mathbb{E}_{X|Z}(nJ_{1n}) = n^{-1}\sum_{j,j'=1}^{n}\gamma(t_j-t_{j'})a\{(t_j-t_{j'})/l_n\}$ and

$$\mathrm{var}_{X|Z}(nJ_{1n}) = n^{-2}\sum_{j_1,j_2,j_1',j_2'=1}^{n}a\{(t_{j_1}-t_{j_1'})/l_n\}a\{(t_{j_2}-t_{j_2'})/l_n\}$$
$$\times\big[\gamma(t_{j_1}-t_{j_2})\gamma(t_{j_1'}-t_{j_2'})+\gamma(t_{j_1}-t_{j_2'})\gamma(t_{j_2}-t_{j_1'})$$
$$+\mathrm{cum}\{X(t_{j_1}),X(t_{j_2}),X(t_{j_1'}),X(t_{j_2'})\}\big].$$

We first show that $|\mathbb{E}_Z\{\mathrm{var}_{X|Z}(nJ_{1n})\}| = o(1)$. Toward this end, we decompose $\sum_{j_1,j_2,j_1',j_2'=1}^{n}$ in $\mathrm{var}_{X|Z}(nJ_{1n})$ into two parts: $N(j_1,j_2,j_1',j_2') := \{(j_1,j_2,j_1',j_2'): j_1\ne j_2\ne j_1'\ne j_2', j_1\ne j_1', j_2\ne j_2\}$ and its complement, and we designate their contributions to $\mathrm{var}_{X|Z}(nJ_{1n})$ as V_{1n} and V_{2n}, respectively. In other words,

$$V_{1n} = n^{-2}\sum_{(j_1,j_2,j_1',j_2')\in N(j_1,j_2,j_1',j_2')}a\{(t_{j_1}-t_{j_1'})/l_n\}a\{(t_{j_2}-t_{j_2'})/l_n\}$$
$$\times\big[\gamma(t_{j_1}-t_{j_2})\gamma(t_{j_1'}-t_{j_2'})+\gamma(t_{j_1}-t_{j_2'})\gamma(t_{j_2}-t_{j_1'})$$
$$+\mathrm{cum}\{X(t_{j_1}),X(t_{j_2}),X(t_{j_1'}),X(t_{j_2'})\}\big]$$
$$= V_{11n}+V_{12n}+V_{13n}.$$

We first analyze V_{11n}. For $(j_1,j_2,j_1',j_2')\in N(j_1,j_2,j_1',j_2')$,

$$\mathbb{E}_Z\big[a\{(t_{j_1}-t_{j_1'})/l_n\}a\{(t_{j_2}-t_{j_2'})/l_n\}\gamma(t_{j_1}-t_{j_2})\gamma(t_{j_1'}-t_{j_2'})\big]$$
$$= \int_{R_0^4}a\{\lambda_n(z_1-z_2)/l_n\}a\{\lambda_n(z_3-z_4)/l_n\}\gamma\{\lambda_n(z_1-z_3)\}$$
$$\times\gamma\{\lambda_n(z_2-z_4)\}\eta(z_1)\eta(z_2)\eta(z_3)\eta(z_4)\,dz_1\,dz_2\,dz_3\,dz_4$$
$$= \lambda_n^{-4}\int_{R_n^4}a\{(z_1-z_2)/l_n\}a\{(z_3-z_4)/l_n\}\gamma\{(z_1-z_3)\}$$
$$\times\gamma\{(z_2-z_4)\}\eta(z_1/\lambda_n)\eta(z_2/\lambda_n)\eta(z_3/\lambda_n)$$
$$\times\eta(z_4/\lambda_n)\,dz_1\,dz_2\,dz_3\,dz_4$$
$$= \lambda_n^{-4}V_{111n}.$$

Because $\eta(\cdot)$ is bounded on R_0 and $a(\cdot)$ has compact support on $[-1,1]$, we can easily show that $|V_{111n}| \le C l_n^2 \lambda_n$ under (7), which implies that $|\mathbb{E}_Z(V_{11n})| \le C\lambda_n l_n^2/n^2$. By a similar argument, $|\mathbb{E}_Z(V_{12n})| \le C\lambda_n l_n^2/n^2$. As for V_{13n}, we have, by (8),

$$\begin{aligned}
&\mathbb{E}_Z\big[a\big\{\big(t_{j_1}-t_{j_1'}\big)/l_n\big\}a\big\{\big(t_{j_2}-t_{j_2'}\big)/l_n\big\} \\
&\qquad \times \operatorname{cum}\big\{X\big(t_{j_1}\big),X\big(t_{j_2}\big),X\big(t_{j_1'}\big),X\big(t_{j_2'}\big)\big\}\big] \\
&= \int_{R_0^4} a\{\lambda_n(z_1-z_2)/l_n\}a\{\lambda_n(z_3-z_4)/l_n\} \\
&\qquad \times C_4\{\lambda_n(z_3-z_1),\lambda_n(z_2-z_1),\lambda_n(z_4-z_1)\} \\
&\qquad \times \eta(z_1)\eta(z_2)\eta(z_3)\eta(z_4)\,dz_1\,dz_2\,dz_3\,dz_4 \\
&= \lambda_n^{-4}\int_{R_n^4} a\{(z_1-z_2)/l_n\}a\{(z_3-z_4)/l_n\} \\
&\qquad \times C_4(z_3-z_1,z_2-z_1,z_4-z_1)\eta(z_1/\lambda_n)\eta(z_2/\lambda_n) \\
&\qquad \times \eta(z_3/\lambda_n)\eta(z_4/\lambda_n)\,dz_1\,dz_2\,dz_3\,dz_4 \\
&= O(\lambda_n^{-3}).
\end{aligned}$$

Thus $|\mathbb{E}_Z(V_{13n})| \le Cn^2/\lambda_n^3$ and $|\mathbb{E}_Z V_{1n}| = o(1)$ under the assumption that $l_n = o(n^{1/2})$. To show $|\mathbb{E}_Z V_{2n}| = o(1)$, we can separate the complement of $N(j_1, j_2, j_1', j_2')$ into several subsets and discuss the contributions from each subset. For example, when $j_1 = j_2 \ne j_1' = j_2'$, the corresponding term in $\operatorname{var}_{X|Z}(nJ_{1n})$ is

$$\begin{aligned}
n^{-2} &\sum_{j_1\ne j_1'=1}^{n} a^2\big\{\big(t_{j_1}-t_{j_1'}\big)/l_n\big\} \\
&\qquad \times \big[\gamma^2(0)+\gamma^2\big(t_{j_1}-t_{j_1'}\big)+C_4\big(0,t_{j_1'}-t_{j_1},t_{j_1'}-t_{j_1}\big)\big]. \qquad \text{(A.9)}
\end{aligned}$$

Using the same argument as before, it is not hard to show that the $\mathcal{L}_1$ moment of (A.9) is $O(l_n/\lambda_n) = o(1)$. The contributions from other subsets can be handled similarly and are of order $o(1)$. This leads to $|\mathbb{E}_Z\{\operatorname{var}_{X|Z}(nJ_{1n})\}| = o(1)$.

We next consider $\operatorname{var}_Z[\mathbb{E}_{X|Z}(nJ_{1n})]$. Define $g_n(t) = \mathbb{E}_Z[\gamma(t_1 - t)a\{(t_1-t)/l_n\}]$, where $t_1 = \lambda_n Z_1$. Let $U_j = \sum_{j'=1}^{j-1}[\gamma(t_j - t_{j'})a\{(t_j - t_{j'})/l_n\} - g_n(t_{j'})]$ for $j = 2,\ldots,n$. Write

$$\begin{aligned}
&\mathbb{E}_{X|Z}(nJ_{1n}) - \mathbb{E}_Z\{\mathbb{E}_{X|Z}(nJ_{1n})\} \\
&\quad = 2n^{-1}\sum_{j=2}^{n} U_j + 2n^{-1}\sum_{j=1}^{n-1}(n-j)\big[g_n(t_j) - \mathbb{E}_Z\{g_n(t_j)\}\big]. \qquad \text{(A.10)}
\end{aligned}$$

Let $\mathcal{F}_j^Z = \sigma(Z_1,\ldots,Z_j)$ for $1 \le j \le n$. Then $\{U_j\}_{j=2}^n$ form martingale differences with respect to $\mathcal{F}_j^Z$. Note that $g_n(t) \le C\lambda_n^{-1}$ uniformly in t. Thus we have that

$$\begin{aligned}
&\mathbb{E}_Z\left(n^{-1}\sum_{j=1}^{n-1}(n-j)\big[g_n(t_j)-\mathbb{E}_Z\{g_n(t_j)\}\big]\right)^2 \\
&\quad \le n^{-2}\sum_{j=1}^{n-1}(n-j)^2\mathbb{E}_Z[g_n^2(t_j)] \\
&\quad = O(n/\lambda_n^2) \qquad \text{(A.11)}
\end{aligned}$$

and that $\mathbb{E}_Z(n^{-1}\sum_{j=2}^n U_j)^2 = n^{-2}\sum_{j=2}^n \mathbb{E}_Z(U_j^2) = n^{-2} \times \sum_{j=2}^n \mathbb{E}_Z\{\mathbb{E}(U_j^2|Z_j)\}$. Conditional on Z_j, U_j is a sum of $(j-1)$ iid random variables. Thus $\mathbb{E}(U_j^2|Z_j) = \sum_{j'=1}^{j-1}\mathbb{E}[[\gamma(t_j - t_{j'})a\{(t_j - t_{j'})/l_n\} - g_n(t_{j'})]^2|Z_j]$. Straightforward calculations show that $|\mathbb{E}[\gamma^2(t_j - t_{j'}) \times a^2\{(t_j - t_{j'})/l_n\}|Z_j]| \le C\lambda_n^{-1}$, which yields $|\mathbb{E}(U_j^2|Z_j)| \le C(j-1)\lambda_n^{-1}$ uniformly in $j = 2,\ldots,n$. Consequently, we obtain $\mathbb{E}_Z(n^{-1} \times \sum_{j=2}^n U_j)^2 = O(\lambda_n^{-1}) = o(1)$. In view of (A.10) and (A.11), we get $\operatorname{var}_Z\{\mathbb{E}_{X|Z}(nJ_{1n})\} = o(1)$, which, in conjunction with $|\mathbb{E}_Z\{\operatorname{var}_{X|Z}(n \times J_{1n})\}| = o(1)$, leads to (A.5).

By (A.8), $\bar{X}_n - \mu = O_p(n^{-1/2})$ under case (a). Furthermore, it is not hard to show that $\mathbb{E}_Z[n^{-2}\sum_{j,j'=1}^n a\{(t_j - t_{j'})/l_n\}] = O(l_n/\lambda_n)$, which implies that $nJ_{3n} = O_p(l_n/n) = o_p(1)$. It remains to show (A.7). Write

$$\begin{aligned}
G_n &:= \mathbb{E}\left(\sum_{j,j'=1}^{n}\{X(t_j)-\mu\}a\{(t_j-t_{j'})/l_n\}\right)^2 \\
&= \sum_{j_1,j_2=1}^{n}\sum_{j_1'=1}^{n}\sum_{j_2'=1}^{n}\mathbb{E}_Z\big[\gamma\big(t_{j_1}-t_{j_2}\big)a\big\{\big(t_{j_1}-t_{j_1'}\big)/l_n\big\} \\
&\qquad \times a\big\{\big(t_{j_2}-t_{j_2'}\big)/l_n\big\}\big].
\end{aligned}$$

Analogous to the analysis of $\mathbb{E}_Z \operatorname{var}_{X|Z}(nJ_{1n})$, we can derive the order of G_n by dividing $\sum_{j_1,j_2,j_1',j_2'=1}^n$ into $N(j_1, j_2, j_1', j_2')$ and its complement and treating them separately. After quite tedious but straightforward calculations, we have that $G_n = O(\lambda_n l_n^2)$, which implies (A.7) because $\bar{X}_n - \mu = O_p(n^{-1/2})$. Thus the proof is complete.

SUPPLEMENTAL MATERIALS

Proofs of theorems, lemmas, and additional tables: Proofs of Theorems 4.1, 4.2, and 5.2, some useful lemmas and their proofs, two tables showing the normalized MSEs and coverages for the case of randomly sampled time points with spherical covariance functions. (DWB-final-supp.pdf)

[Received December 2008. Revised August 2009.]

REFERENCES

Anderson, T. W. (1971), *The Statistical Analysis of Time Series*, New York: Wiley. [220]

Andrews, D. W. K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858. [219-221]

Bhattacharya, R. N., and Rao, R. R. (1986), *Normal Approximations and Asymptotic Expansions*, Malabar, FL: R. E. Krieger Publishing Company. [232]

Brillinger, D. R. (1975), *Time Series: Data Analysis and Theory*, San Francisco: Holden-Day. [221]

Bühlmann, P. (1997), "Sieve Bootstrap for Time Series," *Bernoulli*, 3, 123–148. [232]

Bühlmann, P., and Künsch, H. R. (1999), "Block Length Selection in the Bootstrap for Time Series," *Computational Statistics and Data Analysis*, 31, 295–310. [229]

Carlstein, E. (1986), "The Use of Subseries Values for Estimating the Variance of a General Statistic From a Stationary Time Series," *The Annals of Statistics*, 14, 1171–1179. [218]

Carlstein, E., Do, K.-A., Hall, P., Hesterberg, T., and Künsch, H. (1998), "Matched-Block Bootstrap for Dependent Data," *Bernoulli*, 4, 305–328. [218]

Dietrich, C. R., and Newsam, G. N. (1997), "Fast and Exact Simulation of Stationary Gaussian Processes Through Circular Embedding of the Covariance Matrix," *SIAM Journal on Scientific Computing*, 18, 1088–1107. [220]

Doukhan, P. (1994), *Mixing: Properties and Examples. Lecture Notes in Statistics*, Vol. 85, New York: Springer-Verlag. [221]

Franke, J., and Härdle, W. (1992), "On Bootstrapping Kernel Spectral Estimates," *The Annals of Statistics*, 20, 121–145. [232]

Götze, F., and Künsch, H. (1996), "Second Order Correctness of the Blockwise Bootstrap for Stationary Observations," *The Annals of Statistics*, 24, 1914–1933. [223]

Künsch, H. (1989), "The Jackknife and the Bootstrap for General Stationary Observations," *The Annals of Statistics*, 17, 1217–1241. [218]

Hall, P. (1992), *The Bootstrap and Edgeworth Expansion*, New York: Springer. [221]

Hall, P., and Mammen, E. (1994), "On General Resampling Algorithms and Their Performance in Distribution Estimation," *The Annals of Statistics*, 22, 2011–2030. [221]

Hall, P., Horowitz, J. L., and Jing, B.-Y. (1995), "On Blocking Rules for the Bootstrap With Dependent Data," *Biometrika*, 82, 561–574. [229]

Hampel, F., Ronchetti, E., Rousseeuw, P., and Stahel, W. (1986), *Robust Statistics: The Approach Based on Influence Functions*, New York: Wiley. [221]

Hipel, K. W., and Mcleod, A. I. (1994), *Time Series Modelling of Water Resources and Environmental Systems*, Amsterdam, The Netherlands: Elsevier. [218]

Lahiri, S. N. (1992), "Edgeworth Expansion by 'Moving Block' Bootstrap for Stationary and Nonstationary Data," in *Exploring the Limits of Bootstrap*, eds. R. Lepage and L. Billard, New York: Wiley, pp. 183–214. [223]

——— (1999), "Theoretical Comparisons of Block Bootstrap Methods," *The Annals of Statistics*, 27, 386–404. [222]

——— (2003a), *Resampling Methods for Dependent Data*, New York: Springer. [218,221,222,228,231-233]

——— (2003b), "Central Limit Theorems for Weighted Sums of a Spatial Process Under a Class of Stochastic and Fixed Designs," *Sankhyā*, 65, 356–388. [222,223]

——— (2007), "Asymptotic Expansions for Sums of Block-Variables Under Weak Dependence," *The Annals of Statistics*, 35, 1324–1350. [223]

Lahiri, S. N., and Mukherjee, K. (2004), "Asymptotic Distributions of M-Estimators in a Spatial Regression Model Under Some Fixed and Stochastic Spatial Sampling Designs," *Annals of the Institute of Statistical Mathematics*, 56, 225–250. [223]

Lahiri, S. N., and Zhu, J. (2006), "Resampling Methods for Spatial Regression Models Under a Class of Stochastic Designs," *The Annals of Statistics*, 34, 1774–1813. [218,222-224,227]

Lahiri, S. N., Furukawa, K., and Lee, Y. D. (2007), "A Nonparametric Plug-in Method for Selecting the Optimal Block Length," *Statistical Methodology*, 4, 292–321. [229]

Liu, R. Y. (1988), "Bootstrap Procedures Under Some Non-i.i.d. Models," *The Annals of Statistics*, 16, 1696–1708. [219]

Liu, R. Y., and Singh, K. (1992), "Moving Blocks Jackknife and Bootstrap Capture Weak Dependence," in *Exploring the Limits of Bootstrap*, eds. R. LePage and L. Billard, New York: Wiley, pp. 225–248. [218]

Mammen, E. (1993), "Bootstrap and Wild Bootstrap for High-Dimensional Linear Models," *The Annals of Statistics*, 21, 255–285. [219]

Nordman, D., and Lahiri, S. N. (2004), "On Optimal Spatial Subsample Size for Variance Estimation," *The Annals of Statistics*, 32, 1981–2027. [222]

Nordman, D. J., Lahiri, S. N., and Fridley, B. L. (2007), "Optimal Block Size for Variance Estimation by a Spatial Block Bootstrap Method," *Sankhyā*, 69, 468–493. [222,225]

Paparoditis, E., and Politis, D. N. (2001), "Tapered Block Bootstrap," *Biometrika*, 88, 1105–1119. [218-221,224,225,228,229,231]

——— (2002), "The Tapered Block Bootstrap for General Statistics From Stationary Sequences," *Econometrics Journal*, 5, 131–148. [218,221,225,229]

Parzen, E. (1983), *Time Series Analysis of Irregularly Observed Data*. *Lecture Notes in Statistics*, Vol. 25, New York: Springer. [218]

Politis, D. N. (2003), "The Impact of Bootstrap Methods on Time Series Analysis," *Statistical Science*, 18, 219–230. [220]

Politis, D. N., and Romano, J. P. (1992), "A Circular Block Resampling Procedure for Stationary Data," in *Exploring the Limits of Bootstrap*, eds. R. Lepage and L. Billard, New York: Wiley, pp. 263–270. [218]

——— (1993), "Nonparametric Resampling for Homogeneous Strong Mixing Random Fields," *Journal of Multivariate Analysis*, 47, 301–328. [218]

——— (1994), "The Stationary Bootstrap," *Journal of the American Statistical Association*, 89, 1303–1313. [218]

——— (1995), "Bias-Corrected Nonparametric Spectral Estimation," *Journal of Time Series Analysis*, 16, 67–103. [231]

Politis, D. N., and White, H. (2004), "Automatic Block-Length Selection for the Dependent Bootstrap," *Econometric Reviews*, 23, 53–70. [229,231]

Politis, D. N., Paparoditis, E., and Romano, J. P. (1999), "Resampling Marked Point Processes," in *Multivariate Analysis, Design of Experiments, and Survey Sampling: A Tribute to J. N. Srivastava*, ed. S. Ghosh, New York: Marcel Dekker, Inc., pp. 163–185. [218]

Priestley, M. B. (1981), *Spectral Analysis and Time Series*, Vol. 1, New York: Academic. [220]

Romano, J. L., and Thombs, L. A. (1996), "Inference for Autocorrelations Under Weak Assumptions," *Journal of the American Statistical Association*, 91, 590–600. [231]

Rosenblatt, M. (1984), "Asymptotic Normality, Strong Mixing, and Spectral Density Estimates," *The Annals of Probability*, 12, 1167–1180. [220]

Wu, C. F. J. (1986), "Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis" (with discussion), *The Annals of Statistics*, 14, 1261–1350. [218,219]

Zhu, J., and Lahiri, S. N. (2006), "Bootstrapping the Empirical Distribution Function of a Spatial Process," *Statistical Inference for Stochastic Processes*, 10, 107–145. [218]

Zhurbenko, I. G., and Zuev, N. M. (1975), "On Higher Spectral Densities of Stationary Processes With Mixing," *Ukrainian Mathematical Journal*, 27, 452–446. [221]