Statistica Sinica **20** (2010), 000-000

EXTENDED TAPERED BLOCK BOOTSTRAP

Xiaofeng Shao

University of Illinois at Urbana-Champaign

Abstract: We propose a new block bootstrap procedure for time series, called the extended tapered block bootstrap, to estimate the variance and approximate the sampling distribution of a large class of approximately linear statistics. Our proposal differs from the existing tapered block bootstrap (Paparoditis and Politis (2001, 2002)) in that the tapering is applied to the random weights in the bootstrapped empirical distribution. Under the smooth function model, we obtain asymptotic bias and variance expansions for the variance estimator and establish the consistency of the distribution approximation. The extended tapered block bootstrap has wider applicability than the tapered block bootstrap, while preserving the favorable bias and mean squared error properties of the tapered block bootstrap over the moving block bootstrap. A small simulation study is performed to compare the finite-sample performance of the block-based bootstrap methods.

Key words and phrases: Block bootstrap, empirical measure, influence function, lag window estimator, tapering, variance estimation.

1. Introduction

Since the seminal work of Künsch (1989), the nonparametric block bootstrap methods have received a lot of attention in time series literature. As an important extension of Efron's iid (independent and identically distributed) bootstrap to stationary observations, the moving block bootstrap (MBB) (Künsch (1989), and Liu and Singh (1992)) can be used to approximate the variance and sampling distribution of statistics from a time series. A few variants of the MBB have been proposed, such as the circular block bootstrap (Politis and Romano (1992)), the non-overlapping block bootstrap (Carlstein (1986)), and the stationary bootstrap (Politis and Romano (1994)), among others. For variance estimation in the smooth function model, the MBB and its variants yield the same convergence rate of the mean squared error (MSE), albeit with a different constant in the leading term of the bias and variance expansions; see, e.g., Lahiri (1999, 2003) and Nordman (2009). In an attempt to reduce the bias and MSE, Carlstein, Do, Hall, Hesterberg and Künsch (1998) proposed the so-called matched block bootstrap, which links the bootstrapped blocks using a matching algorithm to achieve a bias reduction, but their method only works for Markovian processes. Later, utilizing a connection between the spectrum estimator and bootstrap variance estimator, Paparoditis and Politis (2001) (PP, hereafter) proposed the tapered block bootstrap (TBB), that involves tapering each (overlapping) block of the series first, then a resampling of those tapered blocks. The TBB offers a superior convergence rate in the bias and MSE compared to the MBB and its variants. The validity of the TBB is shown for the sample mean case in PP (2001), and further extended by PP (2002) to a large class of statistics that are approximately linear. The TBB seems to be the current state-of-the-art block bootstrap method as far as the asymptotic accuracy for variance estimation is concerned.

The main goal of this paper is to widen the scope of the applicability of the TBB by developing a new tapered version of the block bootstrap method, called the extended tapered block bootstrap (ETBB). To motivate our work, we first point out the limitation of the TBB method. Consider a univariate strictly stationary process $\{X_t\}_{t\in\mathbb{Z}}$, and denote by F the distribution function of X_1. Suppose the quantity of interest is $\theta = T(F)$. Given a realization of the process $\mathcal{X}_N = \{X_t\}_{t=1}^N$, 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. An approximately linear statistic $T(\rho_N)$ admits an expansion in a neighborhood of F,

$$T(\rho_N) = T(F) + N^{-1}\sum_{t=1}^{N} IF(X_t; F) + R_N, \tag{1.1}$$

where $IF(x;F)$ is the influence function (Hampel, Ronchetti, Rousseeuw, and Stahel (1986))

$$IF(x;F) = \lim_{\epsilon\downarrow 0}\frac{T((1-\epsilon)F + \epsilon\delta_x) - T(F)}{\epsilon},$$

and R_N is the remainder term. Under some regularity conditions that ensure the negligibility of R_N, $\sigma_N^2 = N\mathrm{Var}\,(\hat{\theta}_N) \approx N^{-1}\mathrm{Var}\,\{\sum_{t=1}^N IF(X_t;F)\}$. In practice, $IF(X_t;F)$ is unknown, but can be replaced by its empirical counterpart $IF(X_t;\rho_N)$, so one can estimate σ_N^2 by applying a block bootstrap procedure to $IF(X_t;\rho_N)$. In fact, in the TBB, PP (2002) proposed to apply tapering to $IF(X_t;\rho_N)$, which is (implicitly) assumed to be known once we observe the data $\mathcal{X}_N$. This is true for a large class of statistics, such as smooth functions of a vector mean, but is not necessarily the case for some other important statistics. A prominent example is $\hat{\theta}_N = \mathrm{median}(X_1,\ldots,X_N)$, which consistently estimates the median of the marginal distribution of X_1, denoted as $\theta = F^{-1}(1/2)$. In this case, we have $IF(x;F) = \{1/2 - \mathbf{1}(x\le\theta)\}/f(\theta)$, where $\mathbf{1}(\cdot)$ is the indicator function and $f(x) = F'(x)$ is the density function of the distribution F. So $IF(X_t;\rho_N) = \{1 - 2\mathbf{1}(X_t \le \hat{\theta}_N)\}/f(\hat{\theta}_N)$ is unknown since $f(\cdot)$ is (typically) unknown in practice. Thus the TBB by PP (2002) is not directly applicable to this setting.

As a remedy, we propose to taper the random weights in the bootstrapped empirical measure in the ETBB. As another natural generalization of the tapering idea to the block bootstrap method, the ETBB turns out to have wider applicability than the TBB. In the sample mean case, the ETBB is identical to the TBB; for the smooth function model, the ETBB variance estimator is not identical to, but is asymptotically equivalent to its TBB counterpart in terms of asymptotic bias and variance. A notable distinction between the ETBB and TBB is that the linearization is implicitly used in the TBB, but not in the ETBB.

The following notation is used throughout the paper. For a column vector $x = (x_1, \ldots, x_q)' \in \mathbb{R}^q$, let $\|x\| = (\sum_{j=1}^q x_j^2)^{1/2}$. Let ξ be a random vector. Write $\xi \in \mathcal{L}^p$ $(p > 0)$ if $\|\xi\|_p := [\mathbb{E}(|\xi|^p)]^{1/p} < \infty$, and let $\|\cdot\| = \|\cdot\|_2$. Denote by $\rightarrow_D$ and $\rightarrow_p$ convergence in distribution and in probability, respectively. The symbols $O_p(1)$ and $o_p(1)$ signify being bounded in probability and convergence to zero in probability respectively. Let $N(\mu, \Sigma)$ be a normal distribution with mean μ and covariance matrix Σ. Denote by $\lfloor a \rfloor$ the integer part of a and $a \vee b = \max(a, b)$, $a \wedge b = \min(a, b)$ for any $a, b \in \mathbb{R}$. The paper is organized as follows. Section 2 introduces the main idea of the ETBB and discusses its connections to the MBB and TBB. Section 3 presents asymptotic results to show the validity of the ETBB. Some simulation findings are reported in Section 4 to corroborate our theoretical results. Section 5 concludes with some discussion.

2. Methodology

To describe the idea of the ETBB, we first introduce the MBB procedure. To perform the MBB, we first specify a block size $l = l_N$. Then $k = k_N = \lfloor N/l \rfloor$ is the number of blocks. For the convenience of presentation, assume $N = kl$. Given the data $\mathcal{X}_N$, we form overlapping blocks $\mathcal{B}_j = \{X_{j+1}, \ldots, X_{j+l}\}$, $j = 0, 1, \ldots, N-l$. When we resample k blocks with replacement from the collection $\{\mathcal{B}_j\}_{j=0}^{N-l}$, say, we get $\mathcal{B}_{S_j}$, $j = 1, \ldots, k$, where $S_1, \ldots, S_k$ are iid uniform random variables on $\{0, \ldots, N-l\}$. Bootstrapped pseudo-observations are obtained by concatenating the resampled blocks together. In other words, the k blocks $(X_1^*, \ldots, X_l^*)$ $(= \mathcal{B}_{S_1})$, $(X_{l+1}^*, \ldots, X_{2l}^*)$ $(= \mathcal{B}_{S_2})$, $(X_{N-l+1}^*, \ldots, X_N^*)$ $(= \mathcal{B}_{S_k})$ are iid with distribution $(N-l+1)^{-1} \sum_{t=0}^{N-l} \delta_{(X_{t+1}, \ldots, X_{t+l})}$. Following Künsch (1989), the bootstrapped empirical measure is

$$\rho_N^* = N^{-1} \sum_{j=1}^{k} \sum_{t=S_j+1}^{S_j+l} \delta_{X_t} = N^{-1} \sum_{t=1}^{N} f_t \delta_{X_t}, \quad \text{where} \quad f_t = \#\{j : t-l \le S_j \le t-1\}.$$

The key idea of the ETBB is to modify f_t to allow the (random) weights to be tapered. Define the data-tapering windows $w_l(\cdot)$ as

$$w_l(h) = w\left(\frac{h - 0.5}{l}\right), \quad h \in \mathbb{N}, \tag{2.1}$$

where the function $w(\cdot)$ satisfies the following conditions.

Assumptions 2.1. The function $w(\cdot) : \mathbb{R} \to [0,1]$ has compact support on $[0,1]$, and $w(t) > 0$ for t in a neighborhood of $1/2$. Further, the function $w(t)$ is symmetric about $t = 0.5$ and nondecreasing for $t \in [0, 1/2]$.

Write $\|w_l\|_1 = \sum_{h=1}^{l} |w_l(h)|$ and $\|w_l\|_2 = (\sum_{h=1}^{l} w_l^2(h))^{1/2}$. Take the tapered random weight assigned to X_t to be

$$\tilde{f}_t = \frac{N}{k\|w_l\|_1} \sum_{h=1}^{l} w_l(h)\#\{j : S_j = t-h\} = \frac{N}{k\|w_l\|_1} \sum_{j=1}^{k} \sum_{h=1}^{l} w_l(h)\mathbf{1}(S_j = t-h),$$

and the bootstrapped empirical measure for the ETBB as

$$\tilde{\rho}_N^* = N^{-1} \sum_{t=1}^{N} \tilde{f}_t \delta_{X_t}. \tag{2.2}$$

Note that $N^{-1} \sum_{t=1}^{N} \tilde{f}_t = 1$, so the bootstrapped empirical measure is a proper probability measure. A detailed check of the TBB procedure in PP (2001) shows that the implied random weights for the TBB coincide with the $\{\tilde{f}_t\}_{t=1}^{N}$ defined above. Note that the TBB is applied to the demeaned data $\{X_t - \bar{X}_N\}_{t=1}^{N}$, where $\bar{X}_N = N^{-1} \sum_{t=1}^{N} X_t$, so the points at which the bootstrapped probability measure concentrates are different from those for the ETBB. For the statistic $\hat{\theta}_N = T(\rho_N)$, its bootstrap version is $\hat{\theta}_N^* = T(\tilde{\rho}_N^*)$. Denote by pr^*, $\mathbb{E}^*$, Var^*, Cov^* the probability, expectation, variance, and covariance conditional on the data $\mathcal{X}_N$. Then the ETBB variance estimator of σ_N^2 is $\tilde{\sigma}_N^2 = M_l(kl)\text{Var}^*(\hat{\theta}_N^*)$, where $M_l = \|w_l\|_1^2/(l\|w_l\|_2^2)$ is the scaling factor needed to account for the increase of the variance due to the tapering in (2.2). Similarly, the distribution of $\sqrt{N}(\hat{\theta}_N - \theta)$ can be approximated by its bootstrap counterparts $\sqrt{klM_l}(\hat{\theta}_N^* - \hat{\theta}_N)$ or $\sqrt{klM_l}(\hat{\theta}_N^* - \mathbb{E}^*\hat{\theta}_N^*)$.

To give a heuristic idea why (2.2) is a proper way of doing tapering, we note that for $l \le t$, $s \le N-l$, $\mathbb{E}^*(\tilde{f}_t) = N/(N-l+1)$ and

$$\begin{aligned}
\text{Cov}^*(\tilde{f}_t, \tilde{f}_s) &= \mathbb{E}(\tilde{f}_t\tilde{f}_s) - \mathbb{E}(\tilde{f}_t)\mathbb{E}(\tilde{f}_s) \\
&= \frac{l^2}{\|w_l\|_1^2} \sum_{h,h'=1}^{l} w_l(h)w_l(h') \sum_{j,j'=1}^{k} \mathbb{E}\Big\{\mathbf{1}(S_j = t-h)\mathbf{1}(S_{j'} = s-h')\Big\} \\
&\quad - \frac{N^2}{(N-l+1)^2} \\
&= \frac{l^2}{\|w_l\|_1^2} \sum_{h,h'=1}^{l} w_l(h)w_l(h')\left\{\frac{(k^2-k)}{(N-l+1)^2} + \frac{k\mathbf{1}(t-s = h-h')}{N-l+1}\right\}
\end{aligned}$$

$$
\begin{aligned}
&-\frac{N^2}{(N-l+1)^2}\\
&=O\Big(\frac{1}{k}\Big)+\frac{Nl}{\|w_l\|_1^2(N-l+1)}\sum_{h,h'=1}^{l}w_l(h)w_l(h')\mathbf{1}(t-s=h-h')\\
&\approx O\Big(\frac{1}{k}\Big)+l\|w_l\|_1^{-2}\sum_{h=1+(t-s)\vee 0}^{l+(t-s)\wedge 0}w_l(h)w_l(h-(t-s))\Big(1+O\Big(\frac{1}{k}\Big)\Big). \quad (2.3)
\end{aligned}
$$

Let $v_l(k)=\sum_{j=1}^{l-|k|}w_l(j)w_l(j+|k|)$. Then $\mathrm{Cov}^*(\tilde f_t,\tilde f_s)=lv_l(|t-s|)/\|w_l\|_1^2+O(1/k)$. Further, we can expand $T(\tilde\rho_N^*)$ in a neighborhood of F, analogous to (1.1), $T(\tilde\rho_N^*)=T(F)+N^{-1}\sum_{t=1}^N IF(X_t;F)\tilde f_t+R_N^*$. Under suitable conditions that imply the negligibility of R_N^*, we have

$$
M_lN\mathrm{Var}^*(\hat\theta_N^*)\approx N^{-1}\sum_{t,s=1}^N IF(X_t;F)IF(X_s;F)\mathrm{Cov}^*(\tilde f_t,\tilde f_s)M_l,
$$

which admits the form of a lag window estimator for the spectral density of the stationary process $\{IF(X_t;F)\}$. In the case of the MBB, $w_l(h)=1$, $h=1,\ldots,l$, $v_l(k)=l-|k|$, $M_l=1$, and $\mathrm{Cov}^*(\tilde f_t,\tilde f_s)\approx\max(1-|t-s|/l,0)$, which corresponds to the Bartlett window (kernel) (Künsch (1989)). In the literature of spectrum estimation (Priestley (1981)), it is well-known that the Bartlett window yields a bias of order $1/l$ and a variance of order l/N. If we use a window that is locally quadratic around zero, such as Parzen's window, then the bias is of (optimal) order $1/l^2$, and the variance is still of order l/N. Equivalently, for the ETBB, if we adopt the data-taper windows $w_l(\cdot)$ (see (2.1)) with $v(\cdot)/v(0)$ being locally quadratic around the origin, then the ETBB variance estimator achieves a bias reduction over its MBB counterpart.

In the sample mean case, $T(F)=\int xdF=\mu$, $T(\rho_N)=\bar X_N$, $IF(x;F)=x-\mu$, $IF(x;\rho_N)=x-\bar X_N$, and the remainder term R_N in (1.1) vanishes. For $\hat\theta_N^*$, $T(\tilde\rho_N^*)=N^{-1}\sum_{t=1}^n\tilde f_tX_t=\bar X_N+N^{-1}\sum_{t=1}^N\tilde f_t(X_t-\bar X_N)$. It is not hard to see that the ETBB variance estimator

$$
\tilde\sigma_N^2=M_lN^{-1}\sum_{t,s=1}^N(X_t-\bar X_N)(X_s-\bar X_N)\mathrm{Cov}^*(\tilde f_t,\tilde f_s)
$$

is identical to its TBB counterpart, where the tapering is applied to $\{X_t-\bar X_N\}$ directly; compare PP (2001).

Remark 2.1. As mentioned in PP (2001), the TBB can be linked to Künsch's (1989) tapered block jackknife in the same way that Efron's bootstrap is linked to Tukey's jackknife. In view of the formulation of the tapered block jackknife in terms of empirical measure (see Künsch (1989, eq. 2.3)), the ETBB is naturally connected to the tapered block jackknife. While both the ETBB and tapered block jackknife can be used for variance estimation, the ETBB offers more by providing an estimate of the sampling distribution of the estimator.

Remark 2.2. The idea of tapering the random weights associated with bootstrapped empirical distribution was mentioned in Künsch (1989) (see Eq. 2.12 therein); this was due to a referee of that paper. The suggested bootstrapped empirical measure takes the form

$$\bar{\rho}_n^* = \left(\sum_{t=1}^n W(\frac{t}{l})\right)^{-1} \sum_{t=1}^n W(\frac{t}{l})\delta_{X_t},$$

where $W(t)$ is a positive stationary process with continuous covariance function $R(t)$, independent of X_t. It seems quite difficult to provide rigorous asymptotic results for this interesting idea, and I am not aware of any developments along this line. In contrast, the form of the random weights is explicitly given in the ETBB, and the extension from f_t to its tapered counterpart $\tilde{f}_t$ has great intuitive appeal. In the next section, we provide an asymptotic justification.

3. Theoretical Validity

We establish here the consistency of the ETBB for both variance estimation and distribution approximation. To make our theoretical results broadly applicable, we consider a multivariate strictly stationary time series $X_t \in \mathbb{R}^m$. Denote its marginal distribution by F^m and its mean by $\mu = \mathbb{E}(X_t)$. The ETBB method described in the previous section can be applied to the multivariate case with only a slight modification of notation. In practice, the statistic of interest could be a functional of m-th marginal distribution of a univariate time series $\{Y_t\}_{t\in\mathbb{Z}}$; thus, with $X_t = (Y_t, \ldots, Y_{t+m-1})'$, the m-th marginal distribution of Y_t is identical to the first marginal of the multivariate series X_t.

Note that for the bootstrapped statistic $T(\tilde{\rho}_N^*)$, we have

$$T(\tilde{\rho}_N^*) = T(F^m) + N^{-1}\sum_{t=1}^N IF(X_t; F^m)\tilde{f}_t + R_N^*.$$

The derivation for the asymptotic bias and variance expansions of our bootstrapped variance estimator $\tilde{\sigma}_N^2$ turns out to be very involved. This difficulty was also mentioned on page 1,231 of Künsch (1989). See Remark 3.1 for more discussions. For this reason, we restrict our attention to the smooth function model $\theta = H(\mu)$, where $H : \mathbb{R}^m \to \mathbb{R}$ is a smooth function. This framework is wide enough to include many statistics of practical interest, such as autocovariance, autocorrelation, the Yule-Walker estimator, and other interesting statistics in time series.

Let $\nabla(x) = (\partial H(x)/\partial x_1, \ldots, \partial H(x)/\partial x_m)'$ and $\nabla = \nabla(\mu)$. Note that in this case, $IF(x; F) = \nabla'(x - \mu)$ and we can write (1.1) as

$$H(\bar{X}_N) = H(\mu) + \nabla' N^{-1} \sum_{t=1}^{N} (X_t - \mu) + R_N. \tag{3.1}$$

Further note that $IF(x; \rho_N) = \nabla(\bar{X}_N)'(x - \bar{X}_N)$, so the tapering in the TBB procedure of PP (2002) is applied to $\nabla(\bar{X}_N)'(X_t - \bar{X}_N)$, $t = 1, \ldots, N$. Let $w * w(t) = \int_{-1}^{1} w(x)w(x + |t|)dx$ be the self-convolution of $w(t)$, and $a(x) = w * w(x)/w * w(0)$. Then the TBB variance estimator of σ_N^2 has a (approximate) closed-form as

$$\nabla(\bar{X}_N)' N^{-1} \sum_{t,t'=1}^{N} (X_t - \bar{X}_N)(X_s - \bar{X}_N)' a(\frac{|t-s|}{l}) \nabla(\bar{X}_N)$$

that can be computed without doing any bootstrapping. In contrast, the tapering in the ETBB is applied to the random weights in the bootstrapped empirical measure, and the resulting bootstrapped statistic is $\hat{\theta}_N^* = H(\bar{X}_N^*)$, where $\bar{X}_N^* = N^{-1} \sum_{t=1}^{N} \tilde{f}_t X_t$. It is not hard to see that the two procedures yield asymptotically equivalent variance estimators under the smooth function model; compare Theorem 3.1 with Theorem 2.1 in PP (2002). Similar to (3.1), for the bootstrapped statistic $T(\rho_N^*) = H(\bar{X}_N^*)$ we have

$$H\left(\bar{X}_N^*\right) = H(\mu) + \nabla'(\bar{X}_N^* - \mu) + R_N^*. \tag{3.2}$$

To state the bias and variance expansions for $\tilde{\sigma}_N^2$, we make the following assumptions.

Assumptions 3.1. The function $H : \mathbb{R}^m \to \mathbb{R}$ is 3-times continuously differentiable and $\max\{|D^v H(x)| : |v| = 3\} \le C(1 + \|x\|^\kappa)$, $x \in \mathbb{R}^m$, for some integer $\kappa \ge 1$.

Denote the i-th component of X_t by $X_{t,i}$.

Assumptions 3.2. For any $(i_1, \ldots, i_s) \in \{1, \ldots, m\}^s$, $2 \le s \le r$,

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

Assumption 3.1 is made to control the magnitude of the remainder term R_N^*; see Lahiri (2003). Assumption 3.2 is common in spectral analysis and can be derived under appropriate moment and mixing conditions (Zhurbenko and Zuev (1975)).

Denote by $\{\alpha(k)\}$ the strong mixing coefficients of the process $\{X_t\}_{t\in\mathbb{Z}}$, and let $\Delta(r;\delta) = 1 + \sum_{k=1}^{\infty} k^{2r-1}\alpha(k)^{\delta/(2r+\delta)}$ for some integer $r \geq 1$ and $\delta > 0$. Let $R_{IF}(k) = \nabla' \operatorname{Cov}(X_0, X_k)\nabla$ and $\sigma_F^2 = \sum_{k=-\infty}^{\infty} R_{IF}(k) > 0$.

Theorem 3.1. (i) *Suppose Assumptions* 2.1, 3.1 *and* 3.2 *hold with* $r = 4$. *Assume* $X_t \in \mathcal{L}^{6+2\kappa+\delta}$ *for some* $\delta > 0$, *and* $\Delta(3+\kappa;1) < \infty$. *If* $w*w$ *is twice continuously differentiable around zero,* $l^{-1} + l/N^{1/3} = o(1)$, *and* $\sum_{k=-\infty}^{\infty} |R_{IF}(k)|k^2 < \infty$, *then*

$$\mathbb{E}(\tilde{\sigma}_N^2) = \sigma_F^2 + B_1 l^{-2} + o(l^{-2}), \tag{3.3}$$

where $B_1 = (1/2)([(w*w)''(0)]/[w*w(0)])\sum_{k=-\infty}^{\infty} k^2 R_{IF}(k)$. (ii) *Suppose Assumptions* 2.1, 3.1 *and* 3.2 *hold with* $r = 8$. *Assume* $X_t \in \mathcal{L}^{12+4\kappa+\delta}$, $l^{-1} + l/N = o(1)$, *and* $\Delta(6+2\kappa;1) < \infty$. *Then*

$$\operatorname{var}(\tilde{\sigma}_N^2) = B_2 \frac{l}{N} + o\Big(\frac{l}{N}\Big), \tag{3.4}$$

where $B_2 = 2\sigma_F^4 \int_{-1}^{1}([(w*w)^2(x)]/[(w*w)^2(0)])dx$.

The statement (3.3) still holds if we replace σ_F^2 by σ_N^2, and $l = o(N^{1/4})$, since $\sigma_N^2 = \sigma_F^2 + O(N^{-1/2})$ under the assumptions of Theorem 3.1 (i). Compared to Theorem 2.1 in PP (2002), obtained under the same bias and variance expansions for the TBB, we require a stronger moment assumption. It can actually be relaxed at the expense of a more stringent assumption on the block size l. We omit the details.

Corollary 3.1. *Under the combined assumptions in Theorem* 3.1 (i) *and* (ii), *the optimal bandwidth* $l^{opt} = (4B_1^2/B_2)^{1/5}N^{1/5}$, *and the corresponding MSE is* $\{(4^{-4/5} + 4^{1/5})B_1^{2/5}B_2^{4/5}\}N^{-4/5}(1+o(1))$.

This result follows from Theorem 3.1 and a straightforward calculation.

Thus the optimal MSE corresponding to the ETBB is $O(N^{-4/5})$, improving upon the $N^{-2/3}$ rate of the MBB. The major improvement is on the bias, which is reduced from the MBB's l^{-1} to l^{-2} since $w*w$ is locally quadratic around zero. Also note that the variance is inflated after tapering by a factor of $(3/2)\int_{-1}^{1}([(w*w)^2(x)]/[(w*w)^2(0)])dx$.

The following theorem states the consistency of the ETBB in terms of approximating the sampling distribution of $\sqrt{N}(\hat{\theta}_N - \theta)$.

Theorem 3.2. *Suppose Assumptions* 2.1 *and* 3.1 *hold. Assume that* $w * w$ *is twice continuously differentiable around zero,* $\sum_{k=-\infty}^{\infty} |R_{IF}(k)|k^2 < \infty$, *and* $l^{-1} + l/N^{1/3} = o(1)$. *Further suppose* $X_t \in \mathcal{L}^{6\vee(3+\kappa)+\delta}$ *for some* $\delta > 0$ *and* $\Delta(\lfloor 2+\kappa/2 \rfloor; 1) < \infty$. *Then*

$$\sup_{x\in\mathbb{R}} \Big| P[\sqrt{N}\{H(\bar{X}_N) - H(\mu)\} \le x] \\ -P^*[\sqrt{klM_l}\{H(\bar{X}_N^*) - \mathbb{E}^*(H(\bar{X}_N^*))\} \le x] \Big| = o_p(1), \tag{3.5}$$

$$\sup_{x\in\mathbb{R}} \Big| P[\sqrt{N}\{H(\bar{X}_N) - H(\mu)\} \le x] \\ -P^*[\sqrt{klM_l}\{H(\bar{X}_N^*) - H(\bar{X}_N)\} \le x] \Big| = o_p(1). \tag{3.6}$$

The proofs of Theorems 3.2 and 3.1 are included in the online supplement, that can be found at `http://www.stat.sinica.edu.tw/statistica`.

Remark 3.1. To indicate the difficulty involved in obtaining analogous results to those of Theorems 3.1 and 3.2 for general statistics $T(F^m)$, we resort to the second order von Mises expansion of $T(\rho_N)$ (von Mises (1947), and Fernholz (1983)). Following the notations in Fernholz (2001), write

$$T(\rho_N) = T(F^m) + N^{-1}\sum_{t=1}^{N} \phi_1(X_t) + \frac{1}{2N^2}\sum_{t,s=1}^{N} \phi_2(X_t, X_s) + \mathrm{Rem}_2.$$

Here $\phi_1(x) = IF(x; F^m)$ is the influence function, that corresponds to the Gâteaux derivative of the functional T. The function ϕ_2 can be defined as

$$\phi_2(x,y) = \frac{d^2}{dsdt} T(F(1-s-t) + t\delta_x + s\delta_y)\Big|_{t=0,s=0}$$

and satisfies $\int \phi_2(x,y)dF^m(x) = \int \phi_2(y,x)dF^m(x) = 0$. Similarly, for the bootstrapped statistic,

$$T(\rho_N^*) = T(F^m) + N^{-1}\sum_{t=1}^{N} \phi_1(X_t)\tilde{f}_t + \frac{1}{2N^2}\sum_{t,s=1}^{N} \phi_2(X_t, X_s)\tilde{f}_t\tilde{f}_s + \mathrm{Rem}_2^*.$$

To find a probabilistic or moment bound for $\mathrm{Var}^*(R_N^*)$, we need to bound $\mathrm{Var}^*(\mathrm{Rem}_2^*)$ and $\mathrm{Var}^*(\sum_{t,s=1}^{N} \phi_2(X_t, X_s)\tilde{f}_t\tilde{f}_s)$ accordingly. It seems hard to come up with easily checked regularity conditions in this general setting, and a case-by-case study might be needed here.

4. Simulation Studies

In this section, we study the finite sample performance of the ETBB compared to the MBB and TBB. In the sample mean case, PP (2001) showed that the TBB outperforms the MBB in terms of the finite-sample MSE of variance estimator and empirical coverage probability of bootstrapped-based confidence interval. Since the ETBB is equivalent to the TBB in the sample mean case, this advantage automatically carries over to the ETBB. Here, we first focus on the comparison of the ETBB with the MBB in the case where the TBB is not applicable. We considered the AR(1) model $X_t = \rho X_{t-1} + \varepsilon_t$, $\varepsilon_t \sim$ iid $N(0,1)$ with $\rho = \pm 0.7$. The estimated quantities are the median and 75% quantile of the marginal distribution of X_1. Two sample sizes $N = 200$ and $1,000$ were considered, but only the results for $N = 200$ are shown as we observe qualitatively similar results for $N = 1,000$. Define the following family of trapezoidal functions as

$$w_c^{\text{trap}}(t) = \begin{cases} \dfrac{t}{c}, & \text{if } t \in [0,c], \\ 1, & \text{if } t \in [c, 1-c], \\ \dfrac{(1-t)}{c}, & \text{if } t \in [1-c, 1]. \end{cases}$$

In our simulation, we took $c = 0.43$, since it was found in PP (2001) that $w_{0.43}^{\text{trap}}$ offers the optimal (theoretical) MSE provided we fix the covariance structure of a time series. For each time series and each block size, we generated $1,000$ ETBB and MBB pseudo-series to estimate σ_N^2 and obtained the bootstrap-based critical values. In every bootstrap repetition, the ETBB and MBB pseudo-series was based on the same randomly selected blocks. Then we repeated this procedure 1,000 times and ploted the empirical MSE and the empirical coverage of nominal 95% symmetric confidence intervals as a function of block size in Figures 1 and 2, respectively.

When $\rho = 0.7$, it is seen from Figure 1 (a), (b) that the ETBB outperformed the MBB for a range of block sizes, although the MBB did better for small (and suboptimal) block sizes. Apparently, the optimal MSE (i.e., the MSE corresponding to the empirical optimal block size) for the ETBB was smaller than that for the MBB, consistent with our theory. For $\rho = -0.7$, Figure 1 (c), (d) shows that the ETBB was superior to the MBB uniformly over the range of block sizes examined. An examination of the empirical coverage probabilities plotted in Figure 2 (a) (b) suggests that for $\rho = 0.7$, the optimal coverage of the ETBB was closer to the nominal level than that of the MBB. For $\rho = -0.7$, the coverage for the block size $l = 1$, which is obviously suboptimal for the purpose of variance estimation, was closest to the nominal level. This might be explained

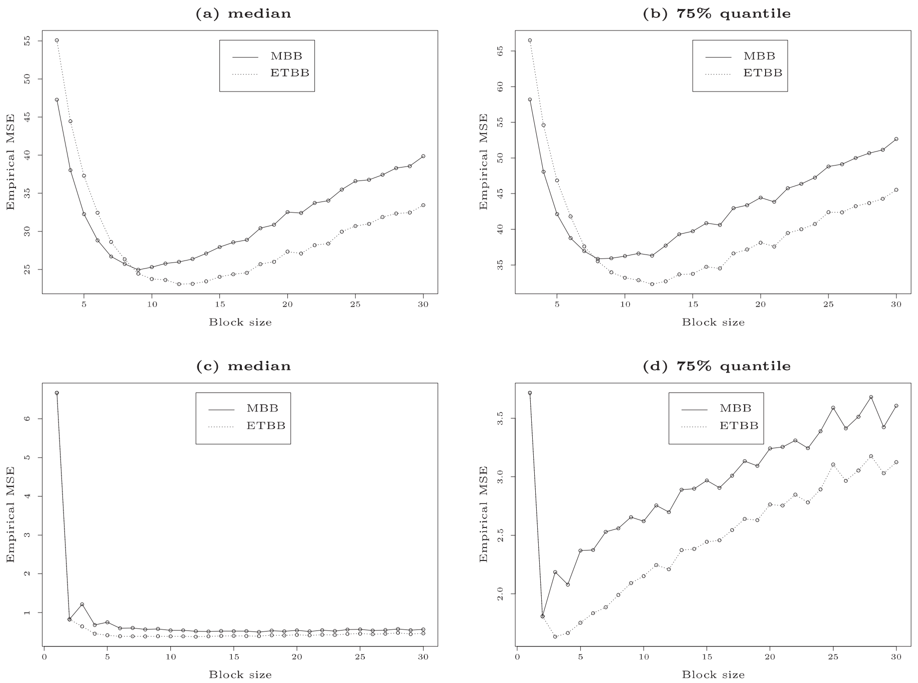

Figure 1. The empirical mean squared error of the MBB and ETBB variance estimators of the variance of the normalized statistics (a) median, $\rho = 0.7$, (b) 75% quantile, $\rho = 0.7$, (c) median, $\rho = -0.7$, and (d) 75% quantile, $\rho = -0.7$.

by the fact that the optimal bandwidth for variance estimation is different from the optimal bandwidth for distribution estimation; see Hall, Horowitz, and Jing (1995). For $3 \leq l \leq 30$, the empirical coverage for the ETBB was fairly close to that of the MBB, and perhaps slightly better than the MBB when $l \geq 10$. Overall, our findings are consistent with those reported in PP (2001, 2002).

For the smooth function model, $\theta = H(\mu)$, where $H(\mu) \neq \mu$, the TBB and ETBB, although asymptotically equivalent, differ in that the ETBB involves no linearization while the TBB does. Thus it would be interesting to examine their finite sample performance in this context. Here we considered the quantity $\gamma(1) = \mathrm{Cov}\,(X_0, X_1)$, which is estimated by its sample analogue $\hat{\gamma}(1) = (N-1)^{-1}\sum_{t=1}^{N-1} X_t X_{t+1} - \{(N-1)^{-1}\sum_{t=1}^{N-1} X_t\}^2$. To put it into the framework of the smooth function model, let $Z_t = (X_t X_{t+1}, X_t)'$. Then $\mu_Z = \mathbb{E}(Z_t) = (\mathbb{E}(X_1X_2), \mathbb{E}(X_1))'$ and $H(\mu_Z) = \mathbb{E}(X_1X_2) - [\mathbb{E}(X_1)]^2$.

From Figure 3, we see that the MBB outperformed both TBB and ETBB when l was small and suboptimal, but the optimal MSE for the MBB was larger than the optimal MSEs for the TBB and ETBB. A possible explanation for this

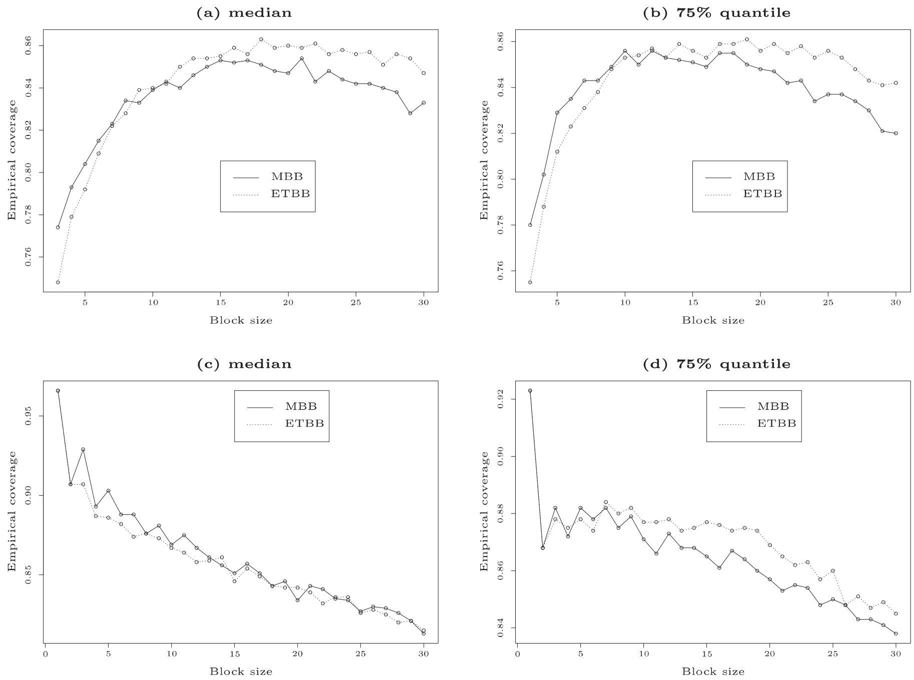

Figure 2. The empirical coverage probability of a two-sided 95% (nominal) confidence interval of (a) median, $\rho = 0.7$, (b) 75% quantile, $\rho = 0.7$, (c) median, $\rho = -0.7$, and (d) 75% quantile, $\rho = -0.7$. Here the bootstrap approximation is based on (3.6).

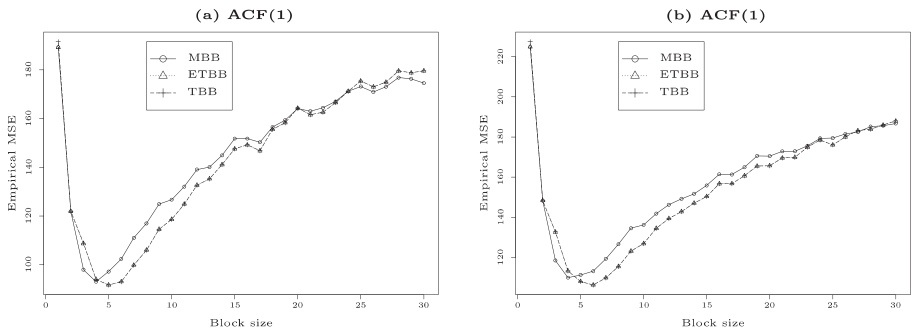

Figure 3. The empirical mean squared error of the MBB, TBB, and ETBB variance estimators of the variance of normalized empirical autocovariance at lag 1 (denoted as ACF(1)). (a) $\rho = 0.7$, (b) $\rho = -0.7$.

advantage of the MBB is that the shape of the taper window cannot fully manifest itself at small block sizes. It is interesting to note that the TBB and ETBB offered almost identical MSEs, which suggests that the linearization involved in the TBB does not cause inaccuracy at the sample size $N = 200$ in this case. Our limited experience with a few other models and smooth functions suggests that the performance of the TBB and ETBB are fairly close. A distinction between the ETBB and the TBB is that no derivative of $H(\cdot)$ needs to be derived in the implementation of the ETBB, while it is needed in the TBB. This could be a disadvantage for the TBB when a closed-form expression for the derivative of $H(\cdot)$ is hard to obtain. In summary, our simulation results provide further support to the claim that the ETBB inherits the advantage of the TBB over the MBB in terms of the MSE of variance estimation, while having broader applicability than the TBB.

5. Discussions

Many issues still merit further research. For the TBB, PP (2001, 2002) remarked on the magnitude of the error term in approximating the sampling distribution of the normalized statistic; compare Theorem 3.2. In view of the connection between the TBB and ETBB, these remarks are expected to apply to the ETBB under appropriate conditions. For the MBB, Götze and Künsch (1996) established second order correctness, i.e., the bootstrap approximation is better than normal approximation. It would be interesting to generalize their results to the TBB and ETBB. An important practical issue is the choice of the block size l. The existing methodologies can be roughly divided into two categories: nonparametric plug-in method (Bühlmann and Künsch (1999), Politis and White (2004), and PP (2001, 2002) among others) and subsampling method (Hall, Horowitz, and Jing (1995)); see Chapter 7 of Lahiri (2003) for a review. For the smooth function model, it is easy to see that the nonparametric plug-in approach works for the ETBB. In particular, the plug-in estimates of B_1 and B_2 can be formed as in PP (2001, 2002) using flat-top lag window estimates. For more general statistics, such as the median, the plug-in approach is still feasible, as noted by Bühlmann and Künsch (1999) for the MBB. In the case of the median, $IF(X_t;\rho_N) = \{1 - 2\mathbf{1}(X_t \leq \hat{\theta}_N)\}/f(\hat{\theta}_N)$, where $\hat{\theta}_N$ is the sample median, is not directly observable. However, the selection of the optimal block size can be based on $\{1 - 2\mathbf{1}(X_t \leq \hat{\theta}_N)\}_{t=1}^N$, since the block length selection is independent of the scale of the data. Further, we note that the subsampling approach of Hall, Horowitz, and Jing (1995) can be easily extended to the ETBB. We leave the finite-sample comparison of the aforementioned block size selection rules to future work. In this article, the theoretical analysis is restricted to the class of smooth function models. In time series analysis, a large class of statistics can be expressed as smooth functionals of empirical processes; see Theorem 4.4 of Lahiri (2003). The consistency of blockwise bootstrapped empirical processes has been studied by Bühlmann (1994), Naik-Nimbalkar and Rajarshi (1994) and

Peligrad (1998) among others, and an extension of their results to the ETBB would be interesting.

Acknowledgements

This research is partially supported by the NSF grants DMS-0804937 and CMG-0724752. Thanks are due to two anonymous referees for their suggestions that improved a previous version of the paper. The author also would like to thank Xuming He for helpful discussions on the von Mises expansion and related asymptotic issues.

References

Brillinger, D. R. (2001). *Time Series: Data Analysis and Theory.* Holden-Day, San-Francisco.

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

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

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

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

Fernholz, L. T. (1983). *Von Mises Calculus for Statistical Functionals.* Lecture notes in Statistics 19, Springer Verlag, New York.

Fernholz, L. T. (2001). On multivariate higher order von Mises expansions. *Metrika* **53**, 123-140.

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

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

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

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

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

Lahiri, S. N. (2003). *Resampling Methods for Dependent Data.* Springer, New York.

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

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

Nordman, D. J. (2009). A note on the stationary bootstrap's variance. *Ann. Statist.* **37**, 359-370.

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

Paparoditis, E. and Politis, D. N. (2002). The tapered block bootstrap for general statistics from stationary sequences. *J. Econometrics* **5**, 131-148.

Peligrad, M. (1998). On the blockwise bootstrap for empirical processes for stationary sequences. *Ann. Probab.* **26**, 877-901.

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

Politis, D. N. and Romano, J. P. (1994). The stationary bootstrap. *J. Amer. Statist. Assoc.* **89**, 1303-1313.

Politis, D. N. and White, H. (2004). Automatic block-length selection for the dependent bootstrap. *Econom. Rev.* **23**, 53-70.

Priestley, M. B. (1981). *Spectral Analysis and Time Series 1.* Academic Press, New York.

von Mises, R. (1947). On the asymptotic distribution of differentiable statistical functions. *Ann. Math. Statist.* **18**, 309-348.

Zhurbenko, I. G. and Zuev, N. M. (1975). On higher spectral densities of stationary processes with mixing. *Ukrain. Matem. Z.* **27** 452-46 (English translation).

Department of Statistics, University of Illinois at Urbana-Champaign, 725 S. Wright St, Champaign, IL 61820, USA.

E-mail: xshao@uiuc.edu

(Received May 2008; accepted January 2009)

Extended Tapered Block Bootstrap

Xiaofeng Shao

University of Illinois at Urbana-Champaign

Supplementary Material

This note contains proofs for Theorems 3.1 and 3.2.

Appendix

In the appendix, $C > 0$ denotes a generic constant that may vary from line to line. Denote by $\mathbb{Z}_+ = \{0, 1, 2 \cdots, \}$. For $v = (v_1, \cdots, v_p)' \in (\mathbb{Z}_+)^p$, $x \in \mathbb{R}^p$, write $x^v = \Pi_{i=1}^p x_i^{v_i}$, $v! = \Pi_{i=1}^p (v_i!)$. For a vector $x = (x_1, \cdots, x_d)' \in \mathbb{R}^d$, let $\|x\|_1 = \sum_{i=1}^d |x_i|$ denote the l^1 norm of x. Write $c_v = D^v H(\mu)/v!$. The following three statements correspond to the equations (4), (5) and (6) in the paper and are needed in our proofs.

$$\begin{aligned} \mathrm{cov}^*(\tilde{f}_t, \tilde{f}_s) &\approx O(1/k) + l\|w_l\|_1^{-2} \sum_{h=1+(t-s)\vee 0}^{l+(t-s)\wedge 0} w_l(h) w_l\{h - (t-s)\} \quad (1.1) \\ &\quad \times\{1 + O(1/k)\}. \\ H(\bar{X}_N) &= H(\mu) + \nabla' N^{-1} \sum_{t=1}^{N} (X_t - \mu) + R_N. \quad (1.2) \\ H\left(\bar{X}_N^*\right) &= H(\mu) + \nabla'(\bar{X}_N^* - \mu) + R_N^*. \quad (1.3) \end{aligned}$$

Proof of Theorem 3.1: (i) We apply a Taylor expansion to $H(\bar{X}_N^*)$ around μ and write $H(\bar{X}_N^*) - H(\mu) = J_{0N} + J_{1N} + J_{2N}$, where

$$\begin{aligned} J_{0N} &= \sum_{\|v\|_1=1} c_v (\bar{X}_N^* - \mu)^v, \quad J_{1N} = \sum_{\|v\|_1=2} c_v (\bar{X}_N^* - \mu)^v, \\ J_{2N} &= 3 \sum_{\|v\|_1=3} (v!)^{-1} (\bar{X}_N^* - \mu)^v \int_0^1 (1-w)^2 D^v H\{\mu + w(\bar{X}_N^* - \mu)\} dw. \end{aligned}$$

In the sequel, we shall show

$$\begin{aligned} \mathbb{E}\{N\mathrm{var}^*(J_{1N})\} &= O(1/N), & (1.4)\\ \mathbb{E}\{N\mathrm{var}^*(J_{2N})\} &= O(1/N^2), & (1.5)\\ \mathbb{E}\{N\mathrm{cov}^*(J_{0N},J_{1N})\} &= O(l/N), & (1.6)\\ \mathbb{E}\{M_l N\mathrm{var}^*(J_{0N})\} &= \sigma_F^2 + B_1 l^{-2} + o(l^{-2}). & (1.7) \end{aligned}$$

If (1.4)-(1.7) hold, then by the Cauchy-Schwarz inequality, we get

$$N\mathbb{E}|\mathrm{cov}^*(J_{0N},J_{2N})| \le N\mathbb{E}^{1/2}\{\mathrm{var}^*(J_{0N})\}\mathbb{E}^{1/2}\{\mathrm{var}^*(J_{2N})\} = O(1/N) = o(l^{-2})$$

and $N\mathbb{E}|\mathrm{cov}^*(J_{1N},J_{2N})| = O(1/N^{3/2}) = o(l^{-2})$ under the assumption that $l = o(N^{1/3})$. Thus the conclusion follows.

To show (1.4), we note that for every $\|v\|_1 = 2$, there exist $v_1, v_2 \in (\mathbb{Z}^+)^m$, $\|v_1\|_1 = 1$, $\|v_2\|_1 = 1$ and $v = v_1 + v_2$. Then we have $(\bar{X}_N^* - \mu)^v = (\bar{X}_N^* - \mu)^{v_1}(\bar{X}_N^* - \mu)^{v_2}$ and

$$\begin{aligned} &\mathrm{var}^*\left\{(\bar{X}_N^*-\mu)^v\right\} = \frac{1}{N^4}\mathrm{var}^*\left[\sum_{t_1,t_2=1}^N (X_{t_1}-\mu)^{v_1}(X_{t_2}-\mu)^{v_2}\{\tilde{f}_{t_1}\tilde{f}_{t_2} - \mathbb{E}(\tilde{f}_{t_1}\tilde{f}_{t_2})\}\right]\\ &\quad = \frac{1}{N^4}\sum_{t_1,t_2,t_3,t_4=1}^N (X_{t_1}-\mu)^{v_1}(X_{t_2}-\mu)^{v_2}(X_{t_3}-\mu)^{v_1}(X_{t_4}-\mu)^{v_2} \qquad (1.8)\\ &\quad \times\mathrm{cov}^*(\tilde{f}_{t_1}\tilde{f}_{t_2},\tilde{f}_{t_3}\tilde{f}_{t_4}). \end{aligned}$$

It is straightforward to show that uniformly in (t_1,t_2,t_3,t_4),

$$\begin{aligned} \mathrm{cov}^*(\tilde{f}_{t_1}\tilde{f}_{t_2},\tilde{f}_{t_3}\tilde{f}_{t_4}) &= \frac{N^4}{k^4\|w_l\|_1^4}\sum_{j_1,j_2,j_3,j_4=1}^k \sum_{h_1,h_2,h_3,h_4=1}^l w_l(h_1)w_l(h_2)w_l(h_3)w_l(h_4)\\ &= \mathrm{cov}^*\{\mathbf{1}(S_{j_1}=t_1-h_1, S_{j_2}=t_2-h_2), \qquad (1.9)\\ &\quad \mathbf{1}(S_{j_3}=t_3-h_3, S_{j_4}=t_4-h_4)\} = O(1). \end{aligned}$$

Hence it follows from Assumption 3.2 (with $r=4$) that

$$\begin{aligned} &\mathbb{E}[N\mathrm{var}^*\{(\bar{X}_N^*-\mu)^v\}] \le \frac{C}{N^3}\sum_{t_1,t_2,t_3,t_4=1}^N |\mathbb{E}\{(X_{t_1}-\mu)^{v_1}(X_{t_2}-\mu)^{v_2}(X_{t_3}-\mu)^{v_1}\\ &\quad (X_{t_4}-\mu)^{v_2}\}| \le \frac{C}{N^3}\sum_{t_1,t_2,t_3,t_4=1}^N \big\{|\mathrm{cov}(X_{t_1}^{v_1},X_{t_2}^{v_2})\mathrm{cov}(X_{t_3}^{v_1},X_{t_4}^{v_2})|\\ &\quad +|\mathrm{cov}(X_{t_1}^{v_1},X_{t_3}^{v_1})\mathrm{cov}(X_{t_2}^{v_2},X_{t_4}^{v_2})| + |\mathrm{cov}(X_{t_1}^{v_1},X_{t_4}^{v_2})\mathrm{cov}(X_{t_2}^{v_2},X_{t_3}^{v_1})|\\ &\quad +|\mathrm{cum}(X_{t_1}^{v_1},X_{t_2}^{v_2},X_{t_3}^{v_1},X_{t_4}^{v_2})|\big\} = O(N^{-1}). \end{aligned}$$

Thus (1.4) holds. Under Assumption 3.1, we have that

$$\mathrm{var}^*(J_{2N}) \leq C\mathbb{E}^*\{\|\bar{X}_N^* - \mu\|^6(1 + \|\mu\|^{2\kappa} + \|\bar{X}_N^* - \mu\|^{2\kappa})\}, \quad (1.10)$$

which implies (1.5) by Lemma 0.1. Next, to show (1.6), we write

$$\begin{aligned}\mathrm{cov}^*(J_{0N}, J_{1N}) &= \sum_{\|v_1\|_1=1} \sum_{\|v_2\|_1=1} \sum_{\|v_3\|_1=1} c_{v_1} c_{v_2+v_3} \\ &\quad \times \mathrm{cov}^*\{(\bar{X}_N^* - \mu)^{v_1}, (\bar{X}_N^* - \mu)^{v_2}(\bar{X}_N^* - \mu)^{v_3}\}.\end{aligned}$$

Note that for any three random variables X, Y, Z, $\mathrm{cov}(X, YZ) = \mathbb{E}[\{X - \mathbb{E}(X)\}\{Y - \mathbb{E}(Y)\}\{Z - \mathbb{E}(Z)\}] + \mathbb{E}(Y)\mathrm{cov}(X, Z) + \mathbb{E}(Z)\mathrm{cov}(X, Y)$. Then for each (v_1, v_2, v_3), we have $\mathrm{cov}^*\{(\bar{X}_N^* - \mu)^{v_1}, (\bar{X}_N^* - \mu)^{v_2}(\bar{X}_N^* - \mu)^{v_3}\} = W_{1N} + W_{2N} + W_{3N}$, where

$$\begin{aligned}W_{1N} &= \mathbb{E}^*[\{(\bar{X}_N^* - \mu)^{v_1} - \mathbb{E}^*((\bar{X}_N^* - \mu)^{v_1})\}\{(\bar{X}_N^* - \mu)^{v_2} - \mathbb{E}^*((\bar{X}_N^* - \mu)^{v_2})\} \\ &\quad \{(\bar{X}_N^* - \mu)^{v_3} - \mathbb{E}^*((\bar{X}_N^* - \mu)^{v_3})\}], \\ W_{2N} &= \mathbb{E}^*(\bar{X}_N^* - \mu)^{v_2}\mathrm{cov}^*\{(\bar{X}_N^* - \mu)^{v_1}, (\bar{X}_N^* - \mu)^{v_3}\}, \\ W_{3N} &= \mathbb{E}^*(\bar{X}_N^* - \mu)^{v_3}\mathrm{cov}^*\{(\bar{X}_N^* - \mu)^{v_1}, (\bar{X}_N^* - \mu)^{v_2}\}.\end{aligned}$$

Write $W_{1N} = N^{-3}\sum_{t_1,t_2,t_3=1}^N (X_{t_1} - \mu)^{v_1}(X_{t_2} - \mu)^{v_2}(X_{t_3} - \mu)^{v_3} CF(t_1, t_2, t_3)$, where

$$\begin{aligned}CF(t_1, t_2, t_3) &:= \mathbb{E}^*[\{\tilde{f}_{t_1} - \mathbb{E}^*(\tilde{f}_{t_1})\}\{\tilde{f}_{t_2} - \mathbb{E}^*(\tilde{f}_{t_2})\}\{\tilde{f}_{t_3} - \mathbb{E}^*(\tilde{f}_{t_3})\}] \\ &= \frac{N^3}{k^3\|w_l\|_1^3} \sum_{j=1}^k \sum_{h_1,h_2,h_3=1}^l w_l(h_1)w_l(h_2)w_l(h_3) \\ &\quad \times \mathbb{E}[\{\mathbf{1}(S_j = t_1 - h_1) - P(S_j = t_1 - h_1)\}\{\mathbf{1}(S_j = t_2 - h_2) \\ &\quad - P(S_j = t_2 - h_2)\}\{\mathbf{1}(S_j = t_3 - h_3) - P(S_j = t_3 - h_3)\}].\end{aligned}$$

It is not hard to see that $|CF(t_1, t_2, t_3)| \leq Cl$ uniformly over (t_1, t_2, t_3). So $|\mathbb{E}(W_{1N})| \leq ClN^{-3}\sum_{t_1,t_2,t_3=1}^N |\mathrm{cum}\{(X_{t_1} - \mu)^{v_1}, (X_{t_2} - \mu)^{v_2}, (X_{t_3} - \mu)^{v_3}\}| = O(l/N^2)$ under Assumption 3.2. Similarly, we have

$$W_{2N} = N^{-3}\sum_{t_1,t_2,t_3=1}^N (X_{t_2} - \mu)^{v_2}(X_{t_1} - \mu)^{v_1}(X_{t_3} - \mu)^{v_3}\mathbb{E}^*(\tilde{f}_{t_2})\mathrm{cov}^*(\tilde{f}_{t_1}, \tilde{f}_{t_3})$$

and $|\mathbb{E}(W_{2N})| \leq CN^{-2}$. The same argument yields $|\mathbb{E}(W_{3N})| = O(N^{-2})$ under Assumption 3.2. Therefore (1.6) holds.

It remains to show (1.7). Let $Y_t = \nabla'(X_t - \mu)$. Then

$$M_l N \mathrm{var}^*(J_{0N}) = \frac{M_l}{N}\sum_{t,s=1}^N Y_t Y_s \mathrm{cov}^*(\tilde{f}_t, \tilde{f}_s). \quad (1.11)$$

In view of (1.1), the above expression is the same as the TBB variance estimator (PP (2001)) except for the mean correction, so (1.7) basically follows from the argument in the proof of Theorem 1 in PP (2001). We omit the details.

(ii). Following the proof of Part (i), the result follows from the following statements:

$$\begin{aligned} \mathrm{var}\{N\mathrm{var}^*(J_{1N})\} &= O(N^{-2}), & (1.12)\\ \mathrm{var}\{N\mathrm{var}^*(J_{2N})\} &= O(N^{-4}), & (1.13)\\ \mathrm{var}\{N\mathrm{var}^*(J_{0N})\} &= B_2 l/N + o(l/N), & (1.14) \end{aligned}$$

since if (1.12)-(1.14) hold, then by the Cauchy-Schwarz inequality and (1.4)-(1.7),

$$\mathrm{var}\{N\mathrm{cov}^*(J_{0N}, J_{1N})\} = O(1/N),\ \ \mathrm{var}\{N\mathrm{cov}^*(J_{0N}, J_{2N})\} = O(N^{-2})$$

and $\mathrm{var}\{N\mathrm{cov}^*(J_{1N}, J_{2N})\} = o(l/N)$.

To show (1.12), we note from (1.8) that for each $v = v_1 + v_2$, $\|v_1\|_1 = 1$, $\|v_2\|_1 = 1$,

$$\begin{aligned} \mathrm{var}[\mathrm{var}^*\{(\bar{X}_N^* - \mu)^v\}] = \frac{1}{N^8} \sum_{t_j=1, j=1,\cdots,8}^{N} & \mathrm{cov}\{(X_{t_1} - \mu)^{v_1}(X_{t_2} - \mu)^{v_2}(X_{t_3} - \mu)^{v_1} \\ & (X_{t_4} - \mu)^{v_2}, (X_{t_5} - \mu)^{v_1}(X_{t_6} - \mu)^{v_2}(X_{t_7} - \mu)^{v_1}(X_{t_8} - \mu)^{v_2}\} \\ & \times \mathrm{cov}^*(\tilde{f}_{t_1}\tilde{f}_{t_2}, \tilde{f}_{t_3}\tilde{f}_{t_4})\mathrm{cov}^*(\tilde{f}_{t_5}\tilde{f}_{t_6}, \tilde{f}_{t_7}\tilde{f}_{t_8}). \end{aligned}$$

By Theorem 2.3.2 in Brillinger (2001), the major summand that involves the covariance of X_ts can be expressed as linear combinations of product of cumulants up to the 8-th order. Assumption 3.2 (with $r = 8$), in conjunction with (1.9), implies that $\mathrm{var}[\mathrm{var}^*\{(\bar{X}_N^* - \mu)^v\}] = O(N^{-4})$, which results in (1.12). According to (1.10) and Lemma 0.1, we have

$$\mathrm{var}\{N\mathrm{var}^*(J_{2N})\} \le CN^2\mathbb{E}\{\mathbb{E}^*\|\bar{X}_N^* - \mu\|^6 + \mathbb{E}^*\|\bar{X}_N^* - \mu\|^{6+2\kappa}\}^2 \le CN^{-4}.$$

Finally, (1.14) follows from (1.11) and the argument in Theorem 2 of PP (2001). The proof is complete.

$\diamondsuit$

Proof of Theorem 3.2: Let $\Sigma_F = \sum_{k=-\infty}^{\infty} \mathrm{cov}(X_0, X_k)$. Since $\sqrt{N}(\bar{X}_N - \mu) \to_D N(0, \Sigma_F)$, we have $\sqrt{N}\{H(\bar{X}_N) - H(\mu)\} \to_D N(0, \sigma_F^2)$ by the delta method.

Let $\Phi(x)$ be the standard normal cumulative distribution function. By Polyā's theorem,

$$\sup_{x\in\mathbb{R}}|P[\sqrt{N}\{H(\bar{X}_N)-H(\mu)\}\leq x]-\Phi(x/\sigma_F)|=o(1) \text{ as } N\to\infty.$$

Then the first assertion follows if we can show that

$$\sup_{x\in\mathbb{R}}|P^*[\sqrt{N}\{H(\bar{X}_N^*)-\mathbb{E}^*(H(\bar{X}_N^*))\}\leq x]-\Phi(x/\sigma_F)|=o_p(1) \text{ as } N\to\infty.$$

Recall the notation $Y_t=\nabla'(X_t-\mu)$. Based on (1.3), we have

$$H(\bar{X}_N^*)-\mathbb{E}^*\{H(\bar{X}_N^*)\}=N^{-1}\sum_{t=1}^{N}Y_t\{\tilde{f}_t-\mathbb{E}^*(\tilde{f}_t)\}+R_N^*-\mathbb{E}^*(R_N^*).$$

Since $\sqrt{N}\mathbb{E}^*|R_N^*|=o_p(1)$, which is to be shown below, it suffices in view of Lemma 4.1 of Lahiri (2003) to show that

$$\sqrt{M_l/(kl)}\sum_{t=1}^{N}Y_t\{\tilde{f}_t-\mathbb{E}^*(\tilde{f}_t)\}\to_D N(0,\sigma_F^2)$$

in probability. Note that

$$(kl)^{-1}\sqrt{M_l}\sum_{t=1}^{N}Y_t\tilde{f}_t=(kl)^{-1}\sum_{j=1}^{k}\sum_{h=1}^{l}w_l(h)\frac{\sqrt{l}}{\|w_l\|_2}Y_{S_j+h},$$

which is identical to the bootstrap sample mean delivered by the TBB applied to the series Y_t (PP (2001)) except for a mean correction. Thus the remaining proof basically follows the argument in the proof of PP's (2001) Theorem 3. We omit the details.

By Slutsky's theorem, the second assertion follows from $\sqrt{N}[\mathbb{E}^*\{H(\bar{X}_N^*)\}-H(\bar{X}_N)]=o_p(1)$. In view of (1.2) and (1.3), it suffices to show that

$$\sqrt{N}\mathbb{E}^*|R_N^*| = o_p(1), \tag{1.15}$$

$$N^{-1/2}\sum_{t=1}^{N}Y_t\{\mathbb{E}^*(\tilde{f}_t)-1\} = o_p(1), \tag{1.16}$$

$$\sqrt{N}R_N = o_p(1). \tag{1.17}$$

The assertion (1.15) is true since

$$\mathbb{E}\{\mathbb{E}^*|J_{1N}|\} \leq C\mathbb{E}\{\mathbb{E}^*\|\bar{X}_N^*-\mu\|^2\}\leq C/N,$$
$$\mathbb{E}\{\mathbb{E}^*|J_{2N}|\} \leq C\mathbb{E}[\mathbb{E}^*\{\|\bar{X}_N^*-\mu\|^3(1+\|\mu\|^\kappa+\|\bar{X}_N^*-\mu\|^\kappa)\}]\leq CN^{-3/2},$$

where we have applied Lemma 0.1. Further, since $\mathbb{E}^*(\tilde{f}_t) = N/(N-l+1)$ when $l \le t \le N-l+1$, and is bounded, (1.16) follows. Finally, a Taylor expansion of $H(\bar{X}_N)$ around μ yields

$$R_N = 2 \sum_{\|v\|_1=2} (v!)^{-1}(\bar{X}_N - \mu)^v \int_0^1 (1-w) D^v H\{\mu + w(\bar{X}_N - \mu)\} dw.$$

Under Assumption 3.1 on $H(\cdot)$, it is straightforward to derive that $\max\{|D^v H(x)| : |v| = 2\} \le C(1 + \|x\|^{\kappa+1})$, so

$$\mathbb{E}(|R_N|) \le C\mathbb{E}\{\|\bar{X}_N - \mu\|^2(1 + \|\mu\|^{\kappa+1} + \|\bar{X}_N - \mu\|^{\kappa+1})\} \le CN^{-1},$$

by Lemma 3.2 of Lahiri (2003). Thus (1.17) holds and this completes the proof. $\diamondsuit$

Lemma **0.1**. *Assume* $X_t \in \mathcal{L}^{r+\delta}$, $\delta > 0$ *for* $r > 2$, $r \in \mathbb{N}$ *and* $\Delta(\lfloor (r+1)/2 \rfloor; 1) < \infty$. *Then* $\mathbb{E}\{\mathbb{E}^*\|\bar{X}_N^* - \mu\|^r\} \le CN^{-r/2}$.

Proof of Lemma 0.1: It suffices to show that for any v_j, $j = 1, \cdots, m$, where v_j is a m-dimensional unit vector with j-th element being 1 and 0 otherwise, $\mathbb{E}\{\mathbb{E}^*|(\bar{X}_N^* - \mu)^{v_j}|^r\} \le CN^{-r/2}$. Denote by $Z_t = Z_t(j) = (X_t - \mu)^{v_j}$. Under our moment and mixing assumptions, by Lemma 3.2 of Lahiri (2003), we have

$$\mathbb{E}\left|\sum_{t=1}^N Z_t\right|^r \le CN^{r/2}. \tag{1.18}$$

Let $H_j := \sum_{h=1}^l w_l(h) \sum_{t=1}^N Z_t\{\mathbf{1}(S_j = t-h) - P(S_j = t-h)\}$. Note that

$$\begin{aligned}
\mathbb{E}^*|(\bar{X}_N^* - \mu)^{v_j}|^r &= N^{-r}\mathbb{E}^*\left|\sum_{t=1}^N Z_t \tilde{f}_t\right|^r \le CN^{-r}\mathbb{E}^*\left|\sum_{j=1}^k H_j\right|^r \\
&\quad + CN^{-r}\left|\sum_{j=1}^k \sum_{h=1}^l w_l(h) \sum_{t=1}^N Z_t P(S_j = t-h)\right|^r \\
&= CN^{-r}(V_1 + V_2).
\end{aligned}$$

It is easy to see that

$$\begin{aligned}
|\mathbb{E}(V_2)| &\le C\mathbb{E}\left|\sum_{t=1}^N Z_t\right|^r \\
&\quad + \frac{C}{N^r}\mathbb{E}\left|\sum_{j=1}^k \sum_{h=1}^l w_l(h) \sum_{t=1}^N Z_t \mathbf{1}(t-h < 0 \text{ or } t-h > N-l)\right|^r
\end{aligned}$$

which is bounded by $CN^{r/2}$ in view of (1.18). Regarding V_1, we apply Burkholder's inequality and get

$$V_1 \leq C\mathbb{E}^* \left| \sum_{j=1}^{k} H_j^2 \right|^{r/2} \leq C \left(\sum_{j=1}^{k} \mathbb{E}^{*2/r} |H_j|^r \right)^{r/2} \leq Ck^{r/2-1} \sum_{j=1}^{k} \mathbb{E}^* |H_j|^r,$$

whereas by (1.18),

$$\begin{aligned} \mathbb{E}^*|H_j|^r \quad &\leq \quad C\mathbb{E}^* \left| \sum_{h=1}^{l} w_l(h) \sum_{t=1}^{N} Z_t \mathbf{1}(S_j = t-h) \right|^r + \frac{C}{N^r} \left| \sum_{h=1}^{l} w_l(h) \sum_{t=h}^{N-l+h} Z_t \right|^r \\ &\leq \quad (N-l+1)^{-1} \sum_{g=0}^{N-l} \left| \sum_{h=1}^{l} w_l(h) Z_{g+h} \right|^r + \frac{C}{N^r} \left| \sum_{h=1}^{l} w_l(h) \sum_{t=h}^{N-l+h} Z_t \right|^r . \end{aligned}$$

By a variant of Lemma 3.2 of Lahiri (2003), $\mathbb{E}|\sum_{h=1}^{l} w_l(h) Z_h|^r \leq Cl^{r/2}$. Thus we can derive $|\mathbb{E}(V_1)| \leq Ck^{r/2} l^{r/2}$ and $\mathbb{E}\{\mathbb{E}^*|(\bar{X}_N^* - \mu)^{v_j}|^r\} \leq CN^{-r/2}$. The conclusion is established.

$\diamondsuit$