Biometrika (2018), **103**, 1, *pp.* 1–18
Printed in Great Britain Advance Access publication on 31 July 2018

Block bootstrap optimality and empirical block selection for sample quantiles with dependent data

BY T. A. KUFFNER
Department of Mathematics and Statistics, Washington University in St. Louis St. Louis, Missouri 63130, U.S.A.
kuffner@wustl.edu

S. M. S. LEE
Department of Statistics and Actuarial Science, The University of Hong Kong Pokfulam Road, Hong Kong
smslee@hku.hk

AND G. A. YOUNG
Department of Mathematics, Imperial College London, London SW7 2AZ U.K.
alastair.young@imperial.ac.uk

SUMMARY

We establish a general theory of optimality for block bootstrap distribution estimation for sample quantiles under mild strong mixing conditions. In contrast to existing results, we study the block bootstrap for varying numbers of blocks. This corresponds to a hybrid between the subsampling bootstrap and the moving block bootstrap, in which the number of blocks is between 1 and the ratio of sample size to block length. The hybrid block bootstrap is shown to give theoretical benefits, and startling improvements in accuracy in distribution estimation in important practical settings. The conclusion that bootstrap samples should be of smaller size than the original sample has significant implications for computational efficiency and scalability of bootstrap methodologies with dependent data. Our main theorem determines the optimal number of blocks and block length to achieve the best possible convergence rate for the block bootstrap distribution estimator for sample quantiles. We propose an intuitive method for empirical selection of the optimal number and length of blocks, and demonstrate its value in a nontrivial example.

Some key words: Hybrid Block Bootstrap; Subsampling; Optimality; Sample Quantile; Weak Dependence.

1. INTRODUCTION

Sample quantile estimation and inference with dependent data is an important problem, with many common applications in statistics, such as time series analysis, Bayesian inference based on Markov chain Monte Carlo samples, and quantile regression, to name a few. Block bootstrap procedures have proven to be effective and popular tools in such problems. However, the optimal choice of block length to achieve the fastest possible convergence rate of the block bootstrap estimator of the distribution of the sample quantile is an open problem. Optimality in this sense is crucial to achieving accurate point estimates, good coverage properties of confidence intervals, as well as scalability and computational efficiency in high dimensions.

While bootstrap theory for the sample quantile problem is fairly well-understood for independent data, there is no existing optimality theory for dependent data. A change from an independent to a dependent context entails a complete revamp of the bootstrap theory, and optimality results known for the independent case do not have a trivial generalisation in the dependent case.

In this paper, we rigorously establish the optimal convergence rate for the block bootstrap estimator for sample quantiles under standard weak dependence conditions of strong mixing, which cover large classes of time series models. We call our approach a hybrid block bootstrap because the fastest convergence rate is achieved by choosing not only the block length, but also the number of blocks, and the optimal choice is in-between using a single block (the subsampling bootstrap) and using the number of blocks prescribed by the moving block bootstrap. The hybrid block bootstrap is seen to achieve remarkable improvements in accuracy for sample quantile distribution estimation compared to the subsampling bootstrap and the moving block bootstrap.

To put our results in a broader context, we mention that optimal block selection is generally an open question for many blockwise statistical procedures with dependent data. The block bootstrap and blockwise empirical likelihood are two common examples. Many recent papers on these topics contain statements to the effect that the sort of optimality theory and methodology we develop in this paper are challenging open questions, in a wide variety of contexts. See, for example, Gregory et al. (2015); Shao & Politis (2013) and Zhang & Shao (2013).

2. BLOCK BOOTSTRAP METHODS

In the Supplementary Material, we provide a review of relevant bootstrap literature. There is little literature on use of block bootstrap methods for the context considered here, which considers a nonsmooth function of dependent data. Sun & Lahiri (2006), Sun (2007) and Sharipov & Wendler (2013) are notable exceptions. Those authors considered block bootstrap approximation for sample quantiles under weak dependence. Sun & Lahiri (2006) established strong consistency of the moving block bootstrap, assuming only a polynomial (strong) mixing rate, for both distribution and variance estimation of the sample quantiles. Sharipov & Wendler (2013) established similar results for the circular block bootstrap utilizing a different set of conditions to take advantage of empirical process theory for the Bahadur-Ghosh representation of the sample quantile. Sun (2007) is particularly relevant to our work, as discussed further below. All of these earlier results assume that the number of blocks tends to infinity with the sample size.

Most recently, Kuffner et al. (2018) established a more general consistency result for a hybrid block bootstrap, for both distribution and variance estimation of sample quantiles. While an exponential mixing rate is assumed, Kuffner et al. (2018) proved weak consistency for *any* number of blocks, $1 \leq b = O(n/\ell)$ as $n \to \infty$, whereas the existing proofs for the moving block bootstrap and circular block bootstrap required that $b \to \infty$, where $b = \lfloor n/\ell \rfloor$. Here, n is the available sample size, and ℓ is the block length. For a real number h, the notation $\lfloor h \rfloor$ is defined as the largest integer $\leq h$, and $\lceil h \rceil$ is the smallest integer $\geq h$. The value of b is the number of resampled blocks to be pasted to form the bootstrap data series. The case $b = 1$ corresponds to the subsampling bootstrap (Politis & Romano, 1994), and the case $b = \lfloor n/\ell \rfloor$ is the standard moving block bootstrap (Künsch, 1989). Therefore, the consistency results in Kuffner et al. (2018) are for a hybrid between the moving block bootstrap and the subsampling bootstrap, and those two extremes are covered by the same theory.

As noted in Kuffner et al. (2018), their theoretical and empirical results suggest that there can be substantial performance improvement, in terms of mean squared errors for both the variance and distribution estimators, when choosing some value of $b > 1$, but less than $\lfloor n/\ell \rfloor$. This suggests the following question: does there exist some optimal choice of the pair (b, ℓ) which provides the best convergence rate for the bootstrap distribution estimator for sample quantiles under weak dependence? We answer that question in the present paper.

Related to the motivation of the present paper is the paper by Sun (2007). She studied the convergence rate of the moving block bootstrap distribution estimator for sample quantiles with dependent data. A strong mixing condition with exponentially decaying mixing coefficients was assumed. An almost sure convergence result was established, and the best rate of convergence was found to be $O(n^{-1/4} \log \log n)$, which is only slightly different from the convergence rate for bootstrap approximation with independent, identically distributed data (Singh, 1981). We consider a weaker polynomial rate condition, which is also slightly weaker than that assumed in Sun & Lahiri (2006). Moreover, we allow the number of blocks to vary, instead of fixing $b = \lfloor n/\ell \rfloor$. Our main theorem establishes the convergence rate of a hybrid bootstrap distribution estimator for sample quantiles. It is a hybrid between the moving block bootstrap ($b = \lfloor n/\ell \rfloor$) and subsampling ($b = 1$) bootstrap. We also apply our theory to the setting of Sun (2007) below.

Aside from our general optimality results being of foundational and practical value, they also indicate that adaptive selection of the number of blocks could yield considerable improvements in convergence rates for block bootstrap distribution estimators. Moreover, Lemma 4 below is of independent interest, as it gives the convergence rate of the block bootstrap distribution estimator, and has bearing on the regular smooth function model. We have included several relevant empirical examples to illustrate the potential gains of optimal choice of the number of blocks, as opposed to using the prescribed value of b for either the subsampling bootstrap ($b = 1$) or the moving block bootstrap ($b = \lfloor n/\ell \rfloor$). In § 6, we give practical guidance as to how to choose (b, ℓ) in a given applied problem, by proposing a procedure for this purpose.

3. PROBLEM SETTING

3.1. *Notation*

Let $\mathbb{Z} \equiv \{0, \pm 1, \pm 2, \ldots\}$ be the set of all integers. Define $\{X_i\}_{i \in \mathbb{Z}}$ to be a doubly-infinite sequence of random variables on the probability space $(\Omega, \mathcal{F}, P)$. The elements of the sequence possess a common distribution function F, and its corresponding quantile function F^{-1}, defined by

$$F^{-1}(p) = \inf\{u : F(u) \geq p\}, \qquad p \in (0, 1).$$

We will study the block bootstrap distribution estimator of a suitably centered and scaled sample quantile. It is assumed throughout that $\{X_i\}_{i \in \mathbb{Z}}$ is a strictly stationary process. The sequence $(X_1, \ldots, X_n)$ denotes a sample of size n from $\{X_i\}_{i \in \mathbb{Z}}$. Denote by $1\{\cdot\}$ the indicator function, so that $1\{A\} = 1$ if event A occurs, $1\{A\} = 0$, otherwise.

3.2. *The Block Bootstrap*

The moving block bootstrap (Künsch, 1989) splits the original sample $(X_1, \ldots, X_n)$ into overlapping blocks of size ℓ, $B_i = (X_i, \ldots, X_{i+\ell-1})$, together constituting a set $\{B_1, \ldots, B_{n-\ell+1}\}$. Let $B_1^*, \ldots, B_b^*$ be a random sample drawn with replacement from the original blocks, where $b = \lfloor n/\ell \rfloor$ is the number of blocks that will be pasted together to form a pseudo-time series. That $B_1^*, \ldots, B_b^*$ is a random sample from $\{B_1, \ldots, B_{n-\ell+1}\}$ means that the sampled blocks are independently and identically distributed according to a discrete uniform distribution on $\{B_1, \ldots, B_{n-\ell+1}\}$. The observations in the ith resampled block, B_i^*, are $X^*_{(i-1)\ell+1}, \ldots, X^*_{i\ell}$, for $1 \leq i \leq b$. Then the moving block bootstrap sample is the concatenation of the resampled blocks, written as

$$\underbrace{X_1^*,\ldots,X_\ell^*}_{B_1^*},\underbrace{X_{\ell+1}^*,\ldots,X_{2\ell}^*}_{B_2^*},\underbrace{X_{2\ell+1}^*,\ldots,X_{(b-1)\ell}^*}_{B_3^*\cdots B_{b-1}^*},\underbrace{X_{(b-1)\ell+1}^*,\ldots,X_{b\ell}^*}_{B_b^*}.$$

Note that this way of constructing the pseudo-time series will reproduce the original dependence structure *asymptotically*.

The subsampling bootstrap (Politis & Romano, 1994), and specifically the overlapping blocks version relevant to the present setting, first splits the original sample into precisely the same overlapping blocks as the moving block bootstrap, each of length ℓ. However, the subsampling bootstrap draws only a single block. A nice property of this procedure is that the original dependence structure in the sample is exactly retained in the single subsample. By contrast, the pseudo-time series constructed by the moving block bootstrap only reproduces the original dependence structure asymptotically.

We define dependence for the sequence of random variables $\{X_i\}_{i\in\mathbb{Z}}$ in terms of the mixing properties of σ-algebras generated by subsets of the sequence which are separated by a distance, in units of time, tending to infinity. For any two sub-σ-algebras of $\mathcal{F}$, say $\mathcal{F}_1$ and $\mathcal{F}_2$, the α-mixing coefficient between $\mathcal{F}_1$ and $\mathcal{F}_2$ is defined to be (Athreya & Lahiri, 2006, Section 16.2.1)

$$\alpha(\mathcal{F}_1,\mathcal{F}_2)\equiv\sup_{A\in\mathcal{F}_1,B\in\mathcal{F}_2}|\mathrm{pr}(A\cap B)-\mathrm{pr}(A)\mathrm{pr}(B)|. \tag{1}$$

Write $\mathcal{F}_k^{k+t}$ for the smallest σ-algebra of subsets of Ω with respect to which X_i, $i=k,\ldots,k+t$, are measurable. Let $\mathcal{F}_{-\infty}^k$ be the smallest σ-algebra which contains the unions of all of the σ-algebras $\mathcal{F}_a^k$ as $a\to-\infty$. That is, $\mathcal{F}_{-\infty}^k$ is a sub-σ-algebra of $\mathcal{F}$, and it is the σ-algebra generated by the random variables $X_a,X_{a+1},\ldots,X_k$ as $a\to-\infty$. Similarly, for $-\infty\le k\le\infty$, let $\mathcal{F}_k^\infty$ be the σ-algebra generated by the random variables $X_{k+1},X_{k+2},\ldots,X_{k+a}$, as $a\to\infty$. The α-mixing coefficient of the sequence $\{X_i\}_{i\in\mathbb{Z}}$ is defined as

$$\alpha(t)\equiv\sup_{k\in\mathbb{Z}}\alpha(\mathcal{F}_{-\infty}^k,\mathcal{F}_{k+t}^\infty),$$

where $\alpha(\cdot,\cdot)$ is defined in (1). If the α-mixing coefficient decays to zero,

$$\lim_{t\to\infty}\alpha(t)=0, \tag{2}$$

then the process $\{X_i\}_{i\in\mathbb{Z}}$ is said to be strongly mixing. The sequence of random variables $\{X_i\}_{i\in\mathbb{Z}}$ is said to be weakly dependent if the process $\{X_i\}_{i\in\mathbb{Z}}$ is strongly mixing, that is if (2) holds.

4. THEORETICAL RESULTS

Assume that $(X_1,\ldots,X_n)$ is a sample of a stationary strong mixing process with mixing coefficient $\alpha(t)$. We assume either a polynomial mixing rate such that $\alpha(t)=O(t^{-\beta})$ for some $\beta\in(5,\infty)$ or an exponential mixing rate such that $\alpha(t)=O(e^{-Ct})$ for some $C>0$. Denote by F the distribution function of X_1 and F_n the empirical distribution function of $(X_1,\ldots,X_n)$.

Define, for $x\in\mathbb{R}$, $\sigma(x)^2=\lim_{n\to\infty}\mathrm{Var}\big\{n^{1/2}F_n(x)\big\}=\sum_{t=-\infty}^{\infty}\mathrm{Cov}\big(1\{X_0\le x\},1\{X_t\le x\}\big)$. Define, for $\ell\in\{1,2,\ldots,n\}$, $b\in\{1,2,\ldots\}$ and $x\in\mathbb{R}$, $J_1,\ldots,J_b$ to be independent random indices uniformly drawn from the set $\{1, \ldots, n-\ell+1\}$,

$$U_i(x) = \ell^{-1} \sum_{t=i}^{i+\ell-1} 1\{X_t \le x\}, \quad i = 1, \ldots, n-\ell+1,$$
$$U_i^*(x) = \ell^{-1} \sum_{t=J_i}^{J_i+\ell-1} 1\{X_t \le x\}, \quad i = 1, \ldots, b,$$

$$\tilde{F}_n(x) = (n-\ell+1)^{-1} \sum_{i=1}^{n-\ell+1} U_i(x), \quad F_n^*(x) = b^{-1} \sum_{i=1}^{b} U_i^*(x).$$

Define, for $p \in (0,1)$,

$$\xi_p = F^{-1}(p), \qquad \hat{\xi}_n = F_n^{-1}(p), \qquad \tilde{\xi}_n = \tilde{F}_n^{-1}(p), \qquad \xi_n^* = F_n^{*-1}(p).$$

Assume that $f = F'$ is defined on a neighbourhood $\mathscr{N}_p$ of ξ_p, with

$$0 < \inf_{x \in \mathscr{N}_p} f(x) \le \sup_{x \in \mathscr{N}_p} f(x) < \infty.$$

THEOREM 1. *Suppose that $n = O(n-\ell)$, $n^{-\frac{4\beta+7}{6(3\beta+5)}} \ell \to \infty$ and $b \ge 1$. Let $x \in \mathbb{R}$ be fixed and $\delta > 0$ be any arbitrarily small constant.*

(i) If polynomial mixing holds with $\beta \in (5, \infty)$ and $\ell = O(b)$, then

$$\begin{aligned}
&\mathrm{pr}\Big\{(b\ell)^{1/2}\big(\xi_n^* - \tilde{\xi}_n\big) \le x \Big| X_1, \ldots, X_n\Big\} \\
&= \mathrm{pr}\Big\{n^{1/2}\big(\hat{\xi}_n - \xi_p\big) \le x\Big\} + O_p\Big\{\ell^{-1} + \ell^{1/2} n^{-1/2} + (b\ell)^{-1/2}\ell^{\delta} \\
&\qquad + n^{-\frac{\beta-1}{2(\beta+1)}+\delta}(b\ell)^{(1-\delta)/4} + n^{-1} b^{\frac{2\beta+1}{4(\beta+1)}-\delta} \ell^{\frac{4\beta+7}{4(\beta+1)}+5\delta}\Big\} \\
&\quad + o_p\Big\{n^{-\frac{\beta-3}{\beta-1}+\delta}(b\ell)^{1/2} + n^{-\frac{3\beta-1}{4(\beta+1)}+\delta}(b\ell)^{1/2} \\
&\qquad + n^{-\frac{\beta(2\beta-3)}{(\beta-1)(2\beta+1)}+\delta} b^{\frac{1}{2}} \ell^{\frac{1}{2}+\frac{2(\beta+3)}{(\beta-1)(2\beta+1)}} + n^{-\frac{4\beta+5}{4(\beta+1)}+\delta} b^{\frac{1}{2}} \ell^{\frac{\beta+2}{\beta+1}} \\
&\qquad + n^{-\frac{2(\beta+1)}{2\beta+1}+\delta} b^{\frac{1}{2}} \ell^{\frac{4\beta+7}{2(2\beta+1)}} + n^{-\frac{4\beta^2+3\beta+1}{2(2\beta+1)(\beta+1)}+\delta} b^{\frac{1}{2}} \ell^{\frac{3\beta+4}{2(2\beta+1)}}\Big\}.
\end{aligned}$$

(ii) If exponential mixing holds with $\alpha(t) = O(e^{-Ct})$ for some $C > 0$, then

$$\begin{aligned}
&\mathrm{pr}\Big\{(b\ell)^{1/2}\big(\xi_n^* - \tilde{\xi}_n\big) \le x \Big| X_1, \ldots, X_n\Big\} \\
&= \mathrm{pr}\Big\{n^{1/2}\big(\hat{\xi}_n - \xi_p\big) \le x\Big\} + O_p\Big\{\ell^{-1} + \ell^{1/2} n^{-1/2} + (b\ell)^{-1/2} \\
&\qquad + n^{-1} b^{\frac{1}{2}-\delta} \ell^{1+5\delta} + n^{-\frac{1}{2}+\delta}(b\ell)^{(1-\delta)/4}\Big\} \\
&\qquad + o_p\Big\{n^{-\frac{3}{4}+\delta}(b\ell)^{1/2} + n^{-1+\delta} b^{\frac{1}{2}} \ell\Big\}.
\end{aligned}$$

We may deduce from Theorem 1 the following two cases.

Case (i) Polynomial mixing with $\beta \in (5,\infty)$ and $\ell = O(b)$.

The convergence rate of the bootstrap distribution estimator is minimised by setting

$$\ell \propto b \propto \begin{cases} n^{\frac{4\beta+7}{6(3\beta+5)}} \log n, & \beta \in \big(5, (7+185^{1/2})/4\big], \\ n^{\frac{\beta-1}{3(\beta+1)}}, & \beta \in \big(7+185^{1/2})/4, \infty\big), \end{cases}$$

which yields, for any $\delta > 0$,

$$\mathrm{pr}\Big\{(b\ell)^{1/2}\big(\xi_n^* - \tilde{\xi}_n\big) \le x \Big| X_1, \ldots, X_n\Big\} - \mathrm{pr}\Big\{n^{1/2}\big(\hat{\xi}_n - \xi_p\big) \le x\Big\}$$
$$= \begin{cases} O_p\left\{n^{-\frac{14\beta^2+\beta-37}{12(3\beta+5)(\beta+1)}+\delta}\right\}, & \beta \in \big(5, (7+185^{1/2})/4\big], \\ O_p\left\{n^{-\frac{\beta-1}{3(\beta+1)}+\delta}\right\}, & \beta \in \big(7+185^{1/2})/4, \infty\big). \end{cases} \tag{3}$$

Note that as $\beta \to \infty$, the optimal orders of ℓ and b approach $n^{1/3}$, which does not depend on unknown parameters and may be taken as a practical reference for empirical choices of ℓ and b. With such choices, that is $\ell \propto b \propto n^{1/3}$, the bootstrap distribution estimator has the convergence rate $O_p\left\{n^{-\frac{\beta-2}{3(\beta+1)}+\delta}\right\}$, for $\beta \in (5,\infty)$ and any $\delta > 0$. The latter convergence rate is slightly slower than that specified in (3), a price to pay for the absence of knowledge of β.

On the other hand, the moving block bootstrap sets $b = \lfloor n/\ell \rfloor$, based on which the optimal ℓ is of order $n^{1/3}$, so that $b \propto n^{2/3}$. The convergence rate of the resulting bootstrap distribution estimator is given, for any $\delta > 0$, by

$$\begin{cases} O_p\left\{n^{-\frac{\beta-5}{2(\beta-1)}+\delta}\right\}, & \beta \in \big(5, 2+17^{1/2}\big], \\ O_p\left\{n^{-\frac{\beta-3}{4(\beta+1)}+\delta}\right\}, & \beta \in \big(2+17^{1/2}, \infty\big), \end{cases}$$

which is markedly slower than that obtained by setting $\ell \propto b \propto n^{1/3}$. Figure 1 compares the optimal convergence rate with those based on $b \propto \ell \propto n^{1/3}$ and $b = \lfloor n/\ell \rfloor \propto n^{2/3}$, respectively. Log error rates for the block bootstrap distribution estimator are plotted against β for the optimal pairs of (b,ℓ). The choice $b = \ell = n^{1/3}$ is optimal under exponential mixing, and it is our recommendation when no information about the exact value of β is available. Thus the discrepancy between the solid and dashed curves shows how ignorance about β affects the error rate. The choice $b = n^{2/3}, \ell = n^{1/3}$ is optimal for Künsch's moving block bootstrap.

Case (ii) Exponential mixing.

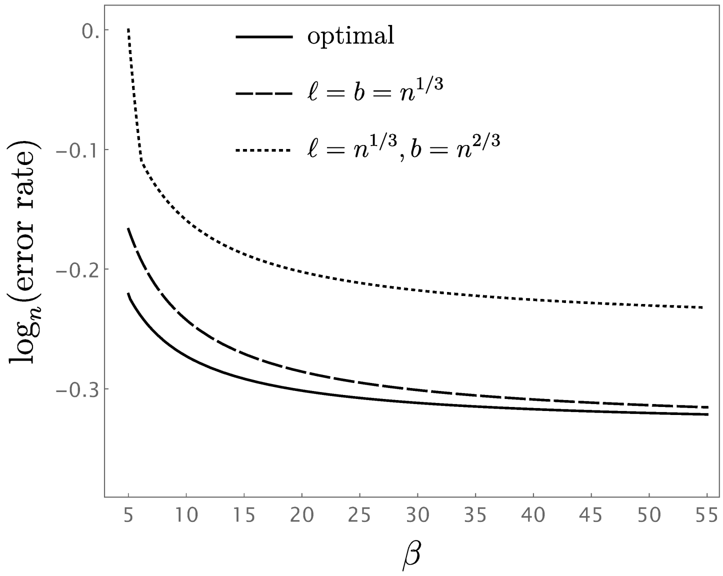
The error rate has an order minimised by setting $\ell \propto b \propto n^{1/3}$, which yields

$$\mathrm{pr}\Big\{(b\ell)^{1/2}\big(\xi_n^* - \tilde{\xi}_n\big) \le x \Big| X_1, \ldots, X_n\Big\} = \mathrm{pr}\Big\{n^{1/2}\big(\hat{\xi}_n - \xi_p\big) \le x\Big\} + O_p\Big(n^{-1/3+\delta}\Big),$$

for any arbitrarily small $\delta > 0$. For moving block bootstrap, the error rate is minimised if ℓ is chosen to have order between $n^{1/4}$ and $n^{1/2}$, yielding an optimal convergence rate of order $O_p\left(n^{-1/4+\delta}\right)$ for any $\delta > 0$. If we set $b = 1$, which amounts to the subsampling method, then the fastest error rate has order $O_p\big(n^{-1/4}\big)$, attained by setting $\ell \propto n^{1/2}$.

Remark 1. The mixing rate could be slower than what we require if the purpose is only to prove that the bootstrap is consistent. For example, Sharipov & Wendler (2013) prove circu-

Fig. 1. Log error rates for the block bootstrap distribution estimator.



lar bootstrap consistency under a very weak condition on the mixing rate. Naturally, stronger conditions are required to investigate higher-order asymptotic properties.

Remark 2. Of independent interest is the result of Lemma 4 in the Appendix, which gives the convergence rate of the block bootstrap distribution estimator for $n^{1/2}\{F_n(x) - F(x)\}$ and has a bearing on the regular smooth function model. Consider the simpler case of exponential mixing. It is easily seen that the convergence rate is minimised at $O_p(n^{-1/3})$, attained by setting $\ell \propto n^{1/3}$ and b having order not smaller than $n^{1/3}$, of which moving block bootstrap is a special case. The subsampling method, $b = 1$, however, has at best a convergence rate of only order $O_p(n^{-1/4})$, attained by setting $\ell \propto n^{1/2}$.

Remark 3. Results on distribution estimation for $n^{1/2}(\hat{\xi}_n - \xi_p)$, embodied in Theorem 1, differ substantially from the regular case in that local estimation of F over a shrinking neighbourhood of size $O_p\{(b\ell)^{-1/2}\}$ around ξ_p incurs an error of order $n^{-1/2}(b\ell)^{1/4}$, which favours a small b and precludes moving block bootstrap from yielding an optimal convergence rate.

Remark 4. The reader may wonder how far the non-bootstrap statistic is from its Gaussian limit. The asymptotic order is given by (A1) in the Appendix. However, to use the Gaussian limit as a potential competing estimator, one must worry about how the variance, which involves the density function, should be estimated optimally. The question of optimality for block bootstrap estimation of the density involves another nonsmooth functional: a kernel density estimator. The density estimation problem is sufficiently different from the sample quantile problem that a separate theory is needed. Our optimality theory for density estimation will be reported elsewhere.

Remark 5. Results analogous to Theorem 1 in the case of independent data have been proved by Sakov & Bickel (2000) and Arcones (2003) for the m out of n bootstrap, which amounts to setting $b = m \to \infty$ and $\ell = 1$ in our block bootstrap procedure. Their proofs build essentially on an Edgeworth expansion for the binomial distribution of $bF_n^*(x)$ to establish asymptotic normality of the bootstrap. With the data strongly mixing and b not necessarily diverging to infinity,

$bF_n^*(x)$ is no longer an expanding sum of independent data. This calls for the critical condition $\ell \to \infty$ and a technically more involved treatment of the cumulant generating function of F_n^* in our proof: see the Appendix for more details.

5. RELEVANCE TO COVERAGE ERROR

Define

$$\hat{G}_n(x) = \mathrm{pr}\left\{(b\ell)^{1/2}(\xi_n^* - \tilde{\xi}_n) \leq x | X_1, \ldots, X_n\right\}$$

and let $\Delta(n, b, \ell)$ be defined by

$$\hat{G}_n(x) = \Phi\left\{xf(\xi_p)/\sigma(\xi_p)\right\} + \Delta(n, b, \ell), \tag{4}$$

where Φ denotes the standard normal distribution function. Our main results in § 4 establish the asymptotic order of $\Delta(n, b, \ell)$ and derive the optimal orders of (b, ℓ) which minimise that order. A level α lower percentile confidence interval for ξ_p is given by

$$[\hat{\xi}_n - n^{-1/2}\hat{G}_n^{-1}(\alpha), \infty).$$

Noting from (4) that

$$\hat{G}_n^{-1}(\alpha) = \Phi^{-1}(\alpha)\sigma(\xi_p)/f(\xi_p) + O_p\left\{\Delta(n, b, \ell)\right\},$$

and using (A1) (Lahiri & Sun, 2009) from the Appendix, we obtain that

$$\begin{aligned}\mathrm{pr}\big\{\xi_p \geq \hat{\xi}_n - n^{-1/2}\hat{G}_n^{-1}(\alpha)\big\} &= \mathrm{pr}\big\{n^{1/2}(\hat{\xi}_n - \xi_p) \leq \Phi^{-1}(\alpha)\sigma(\xi_p)/f(\xi_p)\big\} + O\left\{\Delta(n, b, \ell)\right\} \\ &= \alpha + O\big\{\Delta(n, b, \ell) + n^{-1/2}\big\}.\end{aligned}$$

Since $\Delta(n, b, \ell)$ generally decays at a rate slower than $n^{-1/2}$, which is optimal for independent data, minimising the order of $\Delta(n, b, \ell)$ amounts to minimising the order of the coverage error of the percentile confidence interval.

6. PRACTICAL PROCEDURE FOR SELECTING OPTIMAL (b, ℓ)

Setting $b = \lfloor c_1 n^{1/3} \rfloor$ and $\ell = \lfloor c_2 n^{1/3} \rfloor$, the objective is to find the optimal pair of positive constants (c_1, c_2) which minimise the estimation error of $\hat{G}_n(x)$, or coverage error under some obvious modification of the procedure. Note from (4) and (A1) that

$$\hat{G}_n(x) - G_n(x) = \Delta(n, \lfloor c_1 n^{1/3} \rfloor, \lfloor c_2 n^{1/3} \rfloor) + O(n^{-1/2}). \tag{5}$$

Define, for $c_1, c_2 > 0$ and a fixed $\rho \geq 1$,

$$\delta_n(c_1, c_2) = \big\{\mathbb{E}\big|\Delta\big(n, \lfloor c_1 n^{1/3} \rfloor, \lfloor c_2 n^{1/3} \rfloor\big)\big|^\rho\big\}^{1/\rho}.$$

Then the L_ρ estimation error of $\hat{G}_n(x)$ has the expansion

$$\big\{\mathbb{E}|\hat{G}_n(x) - G_n(x)|^\rho\big\}^{1/\rho} = \delta_n(c_1, c_2) + O(n^{-1/2}). \tag{6}$$

We wish to minimise $\delta_n(c_1, c_2)$ with respect to c_1, c_2.

Let M be a subsample size satisfying $M = o(n)$ and $M \to \infty$. Let $\hat{G}_M^{(j)}(x)$ be constructed analogously to $\hat{G}_n(x)$, with the complete sample $(X_1, \ldots, X_n)$ replaced by the jth block of M

consecutive observations drawn from $(X_1,\ldots,X_n)$, for $j=1,\ldots,n-M+1$. Then we have, analogous to (5), that

$$\hat{G}_M^{(j)}(x)-G_M(x)=\Delta^{(j)}\big(M,\lfloor c_1M^{1/3}\rfloor,\lfloor c_2M^{1/3}\rfloor\big)+O(M^{-1/2}), \tag{7}$$

where $\Delta^{(j)}(\cdot)$ denotes the version of $\Delta(\cdot)$ obtained from the jth subsample. Define

$$Err(c_1,c_2)=(n-M+1)^{-1/\rho}\Big\{\sum_j\big|\hat{G}_M^{(j)}(x)-\hat{G}_n(x)\big|^{\rho}\Big\}^{1/\rho}.$$

Using (A1), (5) and (7), we have

$$\begin{aligned}\hat{G}_M^{(j)}(x)-\hat{G}_n(x)&=G_M(x)+\Delta^{(j)}(M,\lfloor c_1M^{1/3}\rfloor,\lfloor c_2M^{1/3}\rfloor)+O(M^{-1/2})\\&\quad-G_n(x)-\Delta(n,\lfloor c_1n^{1/3}\rfloor,\lfloor c_2n^{1/3}\rfloor)-O(n^{-1/2})\\&=\Delta^{(j)}(M,\lfloor c_1M^{1/3}\rfloor,\lfloor c_2M^{1/3}\rfloor)+O(M^{-1/2}).\end{aligned}$$

It follows that

$$\begin{aligned}Err(c_1,c_2)&=(n-M+1)^{-1/\rho}\Big\{\sum_j\big|\Delta^{(j)}(M,\lfloor c_1M^{1/3}\rfloor,\lfloor c_2M^{1/3}\rfloor)\big|^{\rho}\Big\}^{1/\rho}+O_p(M^{-1/2})\\&=\delta_M(c_1,c_2)\{1+o_p(1)\}+O_p(M^{-1/2}).\end{aligned}$$

If we assume, as is typical, that $\delta_n(c_1,c_2)$ has a leading term of the form $\beta(c_1,c_2)n^{-\gamma}$ $(0<\gamma<1/2)$ for some function $\beta(\cdot)$ independent of n, then $Err(c_1,c_2)$, $\delta_M(c_1,c_2)$ and $\delta_n(c_1,c_2)$ are all minimised at asymptotically the same (c_1,c_2). Thus, an empirical procedure for choosing (c_1,c_2), and hence choosing (b,ℓ), may be based on the minimisation of $Err(c_1,c_2)$.

This procedure constructs the error estimate $Err(c_1,c_2)$ by considering all $n-M+1$ subsamples of M consecutive points drawn from the original data sample, and is therefore computationally expensive. However, the argument supporting minimization of this quantity actually only requires that the number of subsamples used in the construction should grow with sample size n. In practice, therefore, it is reasonable to evaluate the error measure $Err(c_1,c_2)$ using a smaller set of subsamples: in the numerical illustration given below, 20 subsamples, equally spaced along the data series $(X_1,\ldots,X_n)$, are used, allowing rapid evaluation of the error estimate.

7. Examples

To illustrate the benefits of optimally choosing (b,ℓ), we consider three very general examples, the third presented in the Supplementary Material.. For concreteness, we consider $p=1/2$, and simulate the mean squared errors of hybrid block bootstrap estimators of $G_n(u)$ for particular choices of u. The true reference values of $G_n(\cdot)$ are approximated via massive simulation, 5×10^6 replications. For each of the sample sizes $n=200$, $n=500$, and $n=1000$, all entries in the included tables and heat maps are based on 20,000 replications, with 20,000 bootstrap samples used within each replication, unless otherwise stated. For $n=2,000$, the number of replications and bootstrap samples are each $10,000$. For convenience, Table 1 provides some reference values of (b,ℓ) for moving block bootstrap for the sample sizes we consider. This facilitates comparison with the moving block bootstrap choice of $b=\lfloor n/\ell\rfloor$ for a range of values of ℓ. In particular, we give values for ℓ approximately equal to $n^{1/2}$, which is not optimal, $n^{1/3}$, thought to be optimal, $n^{1/4}$, and $n^{1/5}$.

Table 1. *Standard choices of* (b, ℓ) *for different* n*, with the moving block bootstrap choice* $b = \lfloor n/\ell \rfloor$.

	$(b, \ell \approx n^{1/2})$	$(b, \ell \approx n^{1/3})$	$(b, \ell \approx n^{1/4})$	$(b, \ell \approx n^{1/5})$
$n = 200$	$(14, 14)$	$(33, 6)$	$(50, 4)$	$(66, 3)$
$n = 500$	$(22, 22)$	$(62, 8)$	$(100, 5)$	$(125, 4)$
$n = 1,000$	$(31, 32)$	$(100, 10)$	$(166, 6)$	$(250, 4)$
$n = 2,000$	$(44, 45)$	$(153, 13)$	$(285, 7)$	$(400, 5)$

Example 1 (*ARMA(1,1)*). Suppose that the observations are generated according to an ARMA (1,1) model

$$X_t - 0.4X_{(t-1)} = \epsilon_t + 0.3\epsilon_{(t-1)},$$

with ϵ_t independent, identically distributed $N(0,1)$. The strong mixing condition is satisfied with an exponential rate (Lahiri, 2003, Example 6.1). An initial X_0 is sampled according to the marginal distribution, i.e. $X_0 \sim N(0, 1.5833)$, and $\epsilon_0 \sim N(0, 1)$.

With $p = 1/2$, we have $\xi_p = 0$. We simulate the mean squared error in estimation of $G_n(1)$ over a range of (b, ℓ). The true value being estimated was computed, by massive simulation, as described, as $G_n(1) \approx 0.67978$. The heat map in Figure 2 plots mean squared error for $n = 200$, over a grid of values of (b, ℓ). The heat map clearly illustrates the sub-optimality of $b = 1$, the subsampling bootstrap. The minimum mean squared error is 0.00468, with $(b, \ell) = (7, 8)$. By contrast, the minimum mean squared error for the moving block bootstrap is 0.00637, with $(b, \ell) = (33, 6)$, and the subsampling bootstrap, which fixes $b = 1$, has minimum mean squared error of 0.00754, with $\ell = 14$.

Fig. 2. Heat map for the ARMA(1,1) model with $n = 200$.

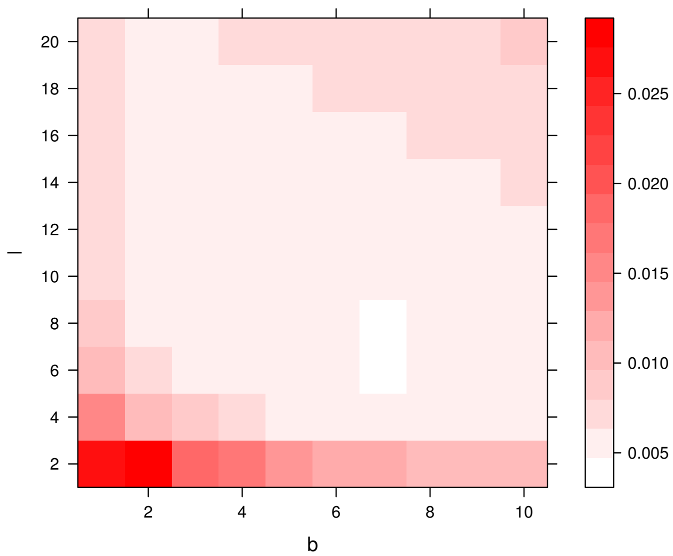


Table 2. *ARMA(1,1) model. Choices of* (b,ℓ) *which minimise the mean squared error for estimating* $G_n(1)$ *for different sample sizes* n.

	(b,ℓ)	MSE
$n=200$	(7,8)	0.00468
$n=500$	(10,10)	0.00250
$n=1,000$	(10,14)	0.00154
$n=2,000$	(12,18)	0.00097

We also compute the values of the pair (b,ℓ) which minimize mean squared error for other sample sizes, $n=500,1000$, and 2000. These results are shown in Table 2. Comparing with Table 1, we note that the mean squared error minimizing pair (b,ℓ) for each n uses an ℓ strictly greater than $n^{1/3}$ and a b much less than $\lfloor n/\ell \rfloor$. Additionally, the mean squared error minimizing value of b is much larger than 1.

The theory says that the hybrid moving block bootstrap has an error rate in estimation of $G_n(1)$ of $n^{-1/3}$, so we should expect the mean squared error to decrease at rate $n^{-2/3}$. In fact, a regression of the logarithm of the mean squared error on $\log(n)$ for the values reported in Table 2 has slope -0.6885, which is not far off $-2/3$. The heat map illustrates that the subsampling and moving block bootstrap choices of (b,ℓ) are suboptimal.

For the current problem, of estimation of the sampling distribution of the sample quantile, there is therefore clear theoretical and practical advantage in using the hybrid block bootstrap, $b\ell < n, b \neq 1$, over the moving block bootstrap. Remark 2 indicates, by contrast, that we might expect to see little difference, in estimation error terms, between the hybrid block bootstrap procedure and moving block bootstrap if, instead, we are interested in estimation of $\mathrm{pr}\big[n^{1/2}\{F_n(x)-F(x)\} \leq y\big]$. This was verified by considering, for all combinations of (b,ℓ), the mean squared error of the estimator $\mathrm{pr}\big[(b\ell)^{1/2}\{F_n^*(x)-\tilde{F}_n(x)\} \leq y \big| X_1,\ldots,X_n\big]$, for $x=0$, so that $F(x)=0.5$, and $y=0.9$, for which the quantity being estimated ≈ 0.89501, for sample size $n=100$. Based on 20,000 replications, with 20,000 bootstrap samples being used in construction of the estimator for each, the minimum mean squared error achieved by moving block bootstrap is 0.00084, with $(b,\ell)=(25,4)$. This is very similar to the overall minimum mean squared error of 0.00082, seen for $(b,\ell)=(18,5)$. The minimum means quared error of the subsampling bootstrap, $b=1$, is 0.00334, substantially larger, when $\ell=7$. This same picture was seen for $n=200$, when, for the same values $x=0, y=0.9$, the true probability being estimated ≈ 0.87781. Simulation shows that the minimum mean squared error of moving block bootstrap is then 0.00108, with $(b,\ell)=(28,7)$, with the same minimum mean squared error for the hybrid block bootstrap, achieved for $(b,\ell)=(30,6)$. Here the subsampling bootstrap yields an optimal mean squared error of 0.00227 when $\ell=8$. These illustrative figures confirm that the hybrid block bootstrap has little advantage over moving block bootstrap in error terms for this problem.

Example 2 (*Nonlinear ARMA(2,3)*). Let $\{X_t\}_{t\in\mathbb{Z}}$ be a sequence from the ARMA(2,3) process

$$X_t - 0.1X_{(t-1)} + 0.3X_{(t-2)} = \epsilon_t + 0.1\epsilon_{(t-1)} + 0.2\epsilon_{(t-2)} - 0.1\epsilon_{(t-3)}.$$

As noted by Lahiri (2003, Example 6.1), such a sequence is strong mixing with exponentially decaying mixing coefficients. To simulate from this model, we initiate by generating X_0, X_{-1} from the marginal $N(0,v^2)$ distribution, which has $v^2=1.0776$, with $\epsilon_0,\epsilon_{-1},\epsilon_{-2}$ independent $N(0,1)$. The nonlinear model we consider is the square transformation of the above ARMA process,

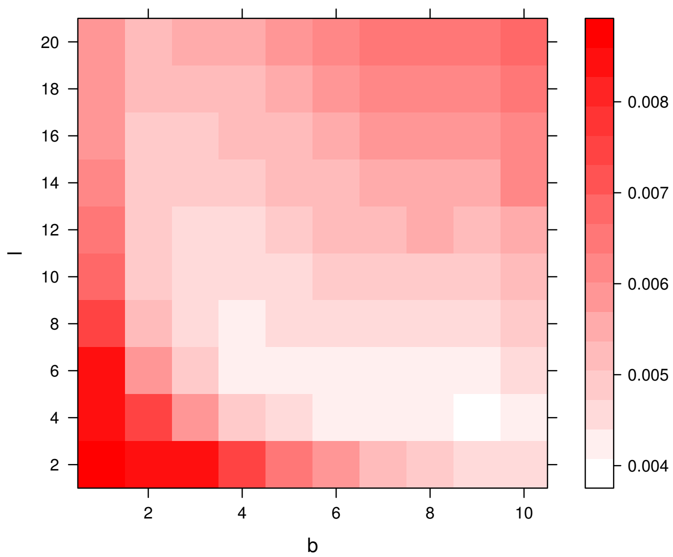
$$Y_t = X_t^2.$$

The square transformation above preserves the strong mixing property and also preserves the mixing rate. Therefore, Y_t is strong mixing with the same exponential rate as X_t. The interested reader is referred to Fan & Yao (2003, p. 69) or Davis & Mikosch (2009, p. 258). As with the previous example, we consider $p = 1/2$, and thus ξ_p satisfies

$$\text{pr}(Y_t = X_t^2 \leq \xi_p) = 1/2,$$

implying $\xi_p = (0.675v)^2$. The simulation approximation to the true value is $G_n(-1.5) \approx 0.09276$.

Fig. 3. Heat map for the nonlinear (squared) ARMA(2,3) model; $n = 200$.



The heat map of Figure 3 shows again that the subsampling and moving block bootstrap choices of (b, ℓ) are suboptimal from the perspective of minimizing mean squared error.

In Figure 4 we display the coverage error of lower percentile confidence intervals, as described in Section 5, of nominal 90% coverage. We observe that there is undercoverage for most choices of (b, ℓ), sometimes very substantial, though there is overcoverage in a few cases. Appropriate choice of (b, ℓ) can yield limits with exactly the required coverage.

As proof of concept of the adaptive procedure for choice of (b, ℓ) described in Section 6, we consider estimation of $G_n(1) \approx 0.80952$, for sample size $n = 512$. We restrict to candidate values $c_1, c_2 \in \{0.5, 0.75, 1.0, 1.5, 2.0\}$, corresponding to adaptive choice of $b, \ell \in \{4, 6, 8, 12, 16\}$. Table 3 shows the mean squared error in estimation of $G_n(1)$ over 2500 replications for each combination of (c_1, c_2). By contrast, the mean squared error obtained by minimization of $Err(c_1, c_2)$ for each replication, using 20 subsamples of size $M = 64$ in construction of this error quantity, was 0.00189. The adaptive method clearly yields a mean squared error that is

Fig. 4. Heatmap for the coverages of 90% lower confidence limits in the nonlinear (squared) ARMA(2,3) model; $n = 200$.

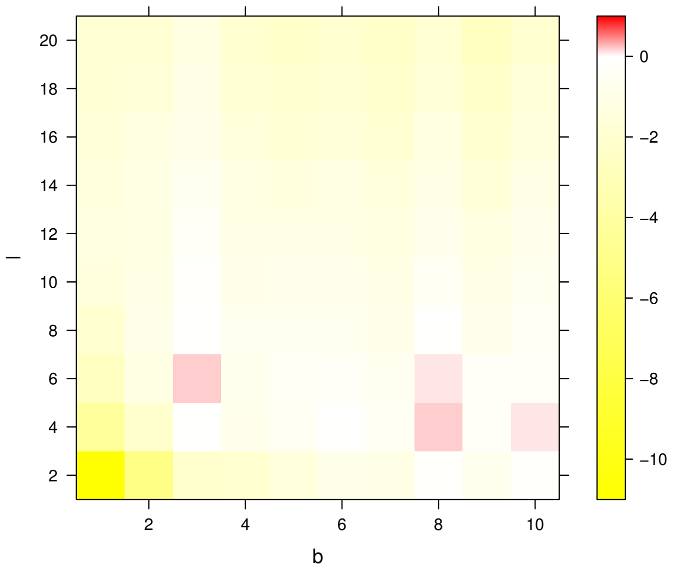


Table 3. *Nonlinear (squared) ARMA(2,3) model: mean squared error in estimation of $G_n(1)$ over 2500 replications, for $b = \lfloor c_1 n^{1/3} \rfloor$ and $\ell = \lfloor c_2 n^{1/3} \rfloor$, $n = 512$. The mean squared error of the adaptive procedure was 0.00189.*

		c_2				
		0.5	0.75	1.0	1.5	2.0
	0.5	0.00164	0.00150	0.00158	0.00189	0.00230
	0.75	0.00143	0.00154	0.00172	0.00220	0.00272
c_1	1.0	0.00144	0.00166	0.00191	0.00250	0.00300
	1.5	0.00161	0.00195	0.00232	0.00297	0.00350
	2.0	0.00179	0.00225	0.00265	0.00335	0.00399

far from optimal in this setting, but outperforms the procedure which fixes b, ℓ to larger values among those being considered.

The adaptive procedure is seen to perform better with increasing sample size. Table 4 provides analagous results for sample size $n = 1728$, for which $G_n(1) \approx 0.81125$. Using $M = 512$ in the minimization of $Err(c_1, c_2)$ over the same range of c_1, c_2, now corresponding to adaptive choice of $b, \ell \in \{6, 9, 12, 18, 24\}$, and again using just 20 subsamples of length M in evaluation of $Err(c_1, c_2)$, the mean squared error of the adaptively chosen estimator over the 2500 replications was observed as 0.00066, much closer to optimal. Further tuning of the adaptive procedure certainly seems worthwhile as a means of providing an effective automatic choice of (b, ℓ) for the hybrid block bootstrap and will be pursued elsewhere.

Table 4. *Nonlinear (squared) ARMA(2,3) model: mean squared error in estimation of $G_n(1)$ over 2500 replications, for $b = \lfloor c_1 n^{1/3} \rfloor$ and $\ell = \lfloor c_2 n^{1/3} \rfloor$, $n = 1728$. The mean squared error of the adaptive procedure was 0.00066.*

		c_2				
		0.5	0.75	1.0	1.5	2.0
	0.5	0.00062	0.00063	0.00072	0.00089	0.00108
	0.75	0.00061	0.00131	0.00082	0.00105	0.00126
c_1	1.0	0.00065	0.00080	0.00094	0.00119	0.00139
	1.5	0.00076	0.00094	0.00112	0.00138	0.00166
	2.0	0.00087	0.00106	0.00125	0.00159	0.00182

In the Supplementary Material we provide a further example involving a process whose mixing coefficients decay at a polynomial rate. This again supports the finding of suboptimality of the choices of (b, ℓ) indicated by the subsampling bootstrap and moving block bootstrap.

Remark 6. Future work will also study the smoothed extended tapered block bootstrap methods of Gregory et al. (2015, 2018), for which only basic consistency results are currently established. Our approach to studying optimal rates is expected to be informative about optimal tuning of such methods, though this latter procedure is complicated by additional tuning parameters.

SUPPLEMENTARY MATERIAL

Supplementary material available at *Biometrika* online includes: a review of the bootstrap literature, derivations of technical results used in the Appendix in the proof of Theorem 1 and a further numerical example.

REFERENCES

ARCONES, M. A. (2003). On the asymptotic accuracy of the bootstrap under arbitrary resampling size. *Annals of the Institute of Statistical Mathematics* **55**, 563–583.

ATHREYA, K. B. & LAHIRI, S. N. (2006). *Measure Theory and Probability Theory*. New York: Springer.

DAVIS, R. A. & MIKOSCH, T. (2009). Probabilistic properties of stochastic volatility models. In *Handbook of Financial Time Series*, T. G. Andersen, R. A. Davis, J.-P. Kreiß & T. Mikosch, eds. Berlin: Springer, pp. 255–267.

FAN, J. & YAO, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. New York: Springer.

FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications, Vol. 2, 2nd ed.* Wiley.

GREGORY, K. B., LAHIRI, S. N. & NORDMAN, D. J. (2015). A smooth block bootstrap for statistical functionals and time series. *Journal of Time Series Analysis* **36**, 442–461.

GREGORY, K. B., LAHIRI, S. N. & NORDMAN, D. J. (2018). A smooth block bootstrap for quantile regression with time series. *Annals of Statistics* **46**, 1138–1166.

KUFFNER, T. A., LEE, S. M. & YOUNG, G. A. (2018). Consistency of block bootstrap for distribution and variance estimation for sample quantiles of weakly dependent sequences. *Australian & New Zealand Journal of Statistics* **60**, 103–114.

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

LAHIRI, S. N. (2003). *Resampling Methods for Dependent Data*. Springer-Verlag.

LAHIRI, S. N. & SUN, S. (2009). A Berry-Esseen theorem for sample quantiles under weak dependence. *Annals of Applied Probability* **19**, 108–126.

POLITIS, D. & ROMANO, J. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *Annals of Statistics* **22**, 2031–2050.

SAKOV, A. & BICKEL, P. J. (2000). An Edgeworth expansion for the m out of n bootstrapped median. *Statistics & Probability Letters* **49**, 217–223.

SHAO, X. & POLITIS, D. N. (2013). Fixed b subsampling and the block bootstrap: improved confidence sets based on p-value calibration. *Journal of the Royal Statistical Society Series B* **75**, 161–184.

SHARIPOV, O. S. & WENDLER, M. (2013). Normal limits, nonnormal limits, and the bootstrap for quantiles of dependent data. *Statistics and Probability Letters* **83**, 1028–1035.

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

SUN, S. (2007). On the accuracy of bootstrapping sample quantiles of strongly mixing sequences. *Journal of the Australian Mathematical Society* **82**, 263–281.

SUN, S. & LAHIRI, S. N. (2006). Bootstrapping the sample quantile of a weakly dependent sequence. *Sankhyā* **68**, 130–166.

ZHANG, X. & SHAO, X. (2013). Fixed-smoothing asymptotics for time series. *Annals of Statistics* **41**, 1329–1349.

APPENDIX: PROOFS

In what follows we denote by C a generic positive constant independent of n. Lahiri & Sun (2009) show under polynomial mixing rates that, for any $x \in \mathbb{R}$,

$$\mathrm{pr}\big\{(n^{1/2}\big(\hat{\xi}_n - \xi_p\big) \le x\big\} = \Phi\big\{xf(\xi_p)/\sigma(\xi_p)\big\} + O(n^{-1/2}). \tag{A1}$$

We first state a lemma which is a special case of Sun and Lahiri's (2006) Lemma 5.3.

LEMMA 1. *Let* $\big\{V_{n,t} : t = 0, \pm1, \pm2, \dots\big\}$ *be a double array of row-wise stationary strong mixing Bernoulli* (p_n) *random variables with* $0 < p_n \le q < 1$ *and mixing coefficients* $\alpha_n(t) = \alpha(t) = O(t^{-\beta})$*, for some fixed* $q \in (0,1)$ *and* $\beta > 0$*. Then, for any positive* $\epsilon_n = o(1)$*,* $n^{-1} \le \delta_n = o(1)$ *and any* $\delta \in (0,1)$*, we have*

$$\begin{aligned}
&\mathrm{pr}\Big\{\Big|\sum_{t=1}^{n}\big(V_{n,t} - p_n\big)\Big| > n\epsilon_n\Big\} \\
&\le C\left(\delta_n^{-1} + \frac{\epsilon_n^2}{p_n + \epsilon_n}\right)\exp\left(-\frac{Cn\delta_n\epsilon_n^2}{p_n + \epsilon_n}\right) + Cn\big(1 + p_n^{\delta}\epsilon_n^{-1}\big)\delta_n^{\beta(1-\delta)}.
\end{aligned}$$

Define, for any $r > 0$, $\mathscr{B}_r(\xi_p) = [\xi_p - r, \xi_p + r]$.

LEMMA 2. *Suppose that* $\alpha(t) = O(t^{-\beta})$ *for some* $\beta > 5$ *and* $n^{-\frac{4\beta+7}{6(3\beta+5)}}\ell \to \infty$*. Then for any arbitrarily small* $\delta > 0$*, the following results hold uniformly over* $\epsilon \in \big[n^{-c_0}, 1\big)$*.*

(i) $$\sup_{x \in \mathscr{B}_\epsilon(\xi_p) \cap \mathscr{N}_p} \big|F_n(x) - F(x)\big| = O_p\left\{n^{-\frac{\beta-1}{2(\beta+1)}+3\delta}\epsilon^{\frac{1}{2(\beta+1)}+\delta}\right\} \text{ for any } c_0 \in (0,3).$$

(ii) $$\sup_{x \in \mathscr{B}_\epsilon(\xi_p) \cap \mathscr{N}_p} \big|\tilde{F}_n(x) - F_n(x)\big| = O_p\left\{n^{-1}\epsilon^{\frac{1}{2(\beta+1)}+\delta}\ell^{\frac{\beta+3}{2(\beta+1)}+3\delta}\right\} \text{ for some } c_0 > 1/2.$$

(iii) $$\sup_{x \in \mathscr{B}_\epsilon(\xi_p) \cap \mathscr{N}_p} \big|F_n(x) - F_n(\xi_p) - F(x) + p\big| = O_p\left\{n^{-\frac{\beta-1}{2(\beta+1)}+\delta}\epsilon^{(1+\delta)/2}\right\} \text{ for any } c_0 \in (0,2).$$

LEMMA 3. *Suppose that* $\alpha(t) = O(t^{-\beta})$ *for some* $\beta > 5$ *and* $n^{-\frac{4\beta+7}{6(3\beta+5)}}\ell \to \infty$*. Then for any arbitrarily small* $\delta > 0$*,*

(i) $$\tilde{\xi}_n = \xi_p + O_p\left\{n^{-1/2} + n^{-\frac{2(\beta+1)}{2\beta+1}+\delta}\ell^{\frac{\beta+3}{2\beta+1}}\right\}.$$

(ii) $$\begin{aligned}\tilde{F}_n\big(\tilde{\xi}_n\big) = p + o_p\big\{&n^{-\frac{\beta-3}{\beta-1}+\delta} + n^{-\frac{\beta(2\beta-3)}{(\beta-1)(2\beta+1)}+\delta}\ell^{\frac{2(\beta+3)}{(\beta-1)(2\beta+1)}} + n^{-\frac{4\beta+5}{4(\beta+1)}+\delta}\ell^{\frac{\beta+3}{2(\beta+1)}} \\ &+ n^{-\frac{2(\beta+1)}{2\beta+1}+\delta}\ell^{\frac{\beta+3}{2\beta+1}}\big\}.\end{aligned}$$

LEMMA 4. *For any arbitrarily small $\delta > 0$ and any compact $\mathscr{K} \subset \mathbb{R}$,*

$$\operatorname{pr}\big[(b\ell)^{1/2}\big\{F_n^*(x) - \tilde{F}_n(x)\big\} \leq y \big| X_1, \ldots, X_n\big] - \Phi\big\{(y/\sigma(x)\big\}$$
$$= \begin{cases} O_p\big\{\ell^{-1} + \ell^{1/2}n^{-1/2} + (b\ell)^{-1/2}\ell^{\delta}\big\} & \text{if } \ell = O(b) \text{ and } \alpha(t) = O(t^{-\beta}) \text{ for some } \beta > 5, \\ O_p\big\{\ell^{-1} + \ell^{1/2}n^{-1/2} + (b\ell)^{-1/2}\big\} & \text{if } \alpha(t) = O(e^{-Ct}), \end{cases}$$

uniformly over $(x, y) \in \mathscr{N}_p \times \mathscr{K}$.

Proof of Lemma 4:

Denote by $\hat{\kappa}_j(x)$ the jth conditional cumulant of $(b\ell)^{1/2}\big\{F_n^*(x) - \tilde{F}_n(x)\big\}$ given $X_1, \ldots, X_n$. It is clear that $\hat{\kappa}_1(x) = 0$.

Define, for $j = 1, 2, \ldots$, $\mathcal{V}_j = (n - \ell + 1)^{-1}\sum_{i=1}^{n-\ell+1}\big\{U_i(x) - F(x)\big\}^j$ and

$$\mathcal{A}_j = \mathbb{E}\Big\{\big\{1\{X_0 \leq x\} - F(x)\big\}\Big[\sum_{|t| \leq \ell-1}\big\{1\{X_t \leq x\} - F(x)\big\}\Big]^{j-1}\Big\}.$$

Then we have, by stationarity and strong mixing properties, $\mathbb{E}\,(\mathcal{V}_j) = \mathbb{E}\big[\{U_1(x) - F(x)\}^j\big] = O\big(\ell^{1-j}\mathcal{A}_j\big)$ and $n\mathrm{Var}\big(\mathcal{V}_j\big) = O\big(\ell^{1-\beta} + \ell^{2-2j}\mathcal{A}_{2j}\big)$.

Consider first the case $\beta < \infty$. Expressing the jth conditional cumulant of $U_1^*(x)$ as a function g_j of $(\mathcal{V}_1, \ldots, \mathcal{V}_j)$, we obtain

$$\begin{aligned} \hat{\kappa}_j(x) &= (b\ell)^{j/2}b^{1-j}g_j(\mathcal{V}_1, \ldots, \mathcal{V}_j) \\ &= (b\ell)^{j/2}b^{1-j}\big[g_j(\mathbb{E}\mathcal{V}_1, \ldots, \mathbb{E}\mathcal{V}_j) + O_p\big\{n^{-1/2}\ell^{(1-\beta)/2} + n^{-1/2}\ell^{1-j}|\mathcal{A}_{2j}|^{1/2}\big\}\big], \end{aligned} \tag{A2}$$

where $g_j(\mathbb{E}\mathcal{V}_1, \ldots, \mathbb{E}\mathcal{V}_j)$ identifies the jth cumulant of $U_1(x) - F(x)$. A comparison with the case of independent data suggests that, for any arbitrarily small $\delta > 0$,

$$g_j(\mathbb{E}\mathcal{V}_1, \ldots, \mathbb{E}\mathcal{V}_j) = O\big(\ell^{-\beta} + \ell^{1-j+\delta}\big). \tag{A3}$$

Noting that $\mathcal{A}_2 = O(1)$ and

$$\mathcal{A}_j = O\Big\{\ell^{j-1}g_j(\mathbb{E}\mathcal{V}_1, \ldots, \mathbb{E}\mathcal{V}_j) + \ell\sum_{2 \leq i \leq j-2}\big|\mathcal{A}_i\mathcal{A}_{j-i}\big|\Big\}, \quad j \geq 3,$$

it can be shown by induction and (A3) that

$$\mathcal{A}_j = O\big\{\ell^{j-1-\beta} + \ell^{\delta} + \ell^{(j-2)/2-(1/2-\delta)1\{j\text{ odd}\}}\big\}, \quad j \geq 3. \tag{A4}$$

It follows from (A2), (A3) and (A4) that

$$\begin{aligned} \hat{\kappa}_2(x) &= \ell g_2(\mathcal{V}_1, \mathcal{V}_2) = \ell(\mathcal{V}_2 - \mathcal{V}_1^2) \\ &= \sum_{1 \leq |t| \leq \ell-1}(1 - |t|/\ell)\operatorname{Cov}\big(1\{X_0 \leq x\}, 1\{X_t \leq x\}\big) + O_p\big\{n^{-1/2}(\ell^{1+(1-\beta)/2} + |\mathcal{A}_4|^{1/2})\big\} \\ &= \sigma(x)^2 + O(\ell^{-1}) + O_p\big\{n^{-1/2}\ell^{(3-\beta)/2} + n^{-1/2}\ell^{1/2}\big\} \\ &= \sigma(x)^2 + O_p\big(\ell^{-1} + n^{-1/2}\ell^{1/2}\big) \end{aligned} \tag{A5}$$

and, for $j \geq 3$ and $\ell = O(b)$,

$$\begin{aligned} \hat{\kappa}_j(x) &= (b\ell)^{j/2}b^{1-j}\big[g_j(\mathbb{E}\mathcal{V}_1, \ldots, \mathbb{E}\mathcal{V}_j) + O_p\big\{n^{-1/2}\ell^{(1-\beta)/2} + n^{-1/2}\ell^{1-j}|\mathcal{A}_{2j}|^{1/2}\big\}\big] \\ &= b^{-(j-2)/2} \times O_p\big\{\ell^{j/2-\beta} + \ell^{1-j/2+\delta} + n^{-1/2}\ell^{(j+1-\beta)/2} + n^{-1/2}\ell^{1/2}\big\} \\ &= O_p\big\{b^{-1/2}\ell^{-1/2+\delta} + n^{-1/2}\ell^{(3-\beta)/2} + n^{-1/2}b^{-1/2}\ell^{1/2}\big\}. \end{aligned} \tag{A6}$$

Without imposing the condition $\ell = O(b)$, the above arguments can similarly be applied to the case of exponential mixing rates to establish (A5) and a stronger version of (A6), with $\delta = 0$ and $\beta = \infty$.

Following Arcones (2003), application of Esseen's lemma (Feller, 1971, Lemma XVI.4.2) to polygonal approximations of the conditional distribution function of $(b\ell)^{1/2}\big\{F_n^*(x) - \tilde{F}_n(x)\big\}$ and $\Phi\big\{\,\cdot\,/\sigma(x)\big\}$ yields, for any arbitrarily large $C' > 0$,

$$\sup_{(x,y)\in\mathscr{N}_p\times\mathscr{K}} \Big|\mathrm{pr}\Big[(b\ell)^{1/2}\big\{F_n^*(x) - \tilde{F}_n(x)\big\} \le y\Big|X_1,\ldots,X_n\Big] - \Phi\big\{y/\sigma(x)\big\}\Big|$$
$$\le CC'^{-1}(b\ell)^{-1/2} + C\int_{-C'\sqrt{b\ell}}^{C'\sqrt{b\ell}} |t|^{-1}e^{-t^2/2}\big|e^{\hat{\kappa}_x^*(t)+t^2/2} - 1\big|\left|\frac{\sin\left\{2^{-1}\sigma(x)^{-1}(b\ell)^{-1/2}t\right\}}{2^{-1}\sigma(x)^{-1}(b\ell)^{-1/2}t}\right|dt,$$

where $\hat{\kappa}_x^*(t)$ denotes the conditional characteristic function of $(b\ell)^{1/2}\big\{F_n^*(x) - \tilde{F}_n(x)\big\}/\sigma(x)$. Lemma 4 then follows by bounding $\hat{\kappa}_x^*(t) + t^2/2$ using (A5) and (A6) under polynomial mixing, or using (A5) and the stronger version of (A6) under exponential mixing.

Proof of Theorem 1:

Consider first the case $\beta < \infty$. We have, by Lemmas 2, 3 and Taylor expansion of F about $\tilde{\xi}_n$,

$$\begin{aligned}
&p - \tilde{F}_n\big\{\tilde{\xi}_n + (b\ell)^{-1/2}x\big\}\\
&= \Big\{p - \tilde{F}_n(\tilde{\xi}_n)\Big\} + \Big[\tilde{F}_n(\tilde{\xi}_n) - \tilde{F}_n\big\{\tilde{\xi}_n + (b\ell)^{-1/2}x\big\}\Big]\\
&= F_n(\tilde{\xi}_n) - F_n\big\{\tilde{\xi}_n + (b\ell)^{-1/2}x\big\} + o_p\Big\{n^{-\frac{\beta-3}{\beta-1}+\delta} + n^{-\frac{\beta(2\beta-3)}{(\beta-1)(2\beta+1)}+\delta}\ell^{\frac{2(\beta+3)}{(\beta-1)(2\beta+1)}}\\
&\quad + n^{-\frac{4\beta+5}{4(\beta+1)}+\delta}\ell^{\frac{\beta+3}{2(\beta+1)}} + n^{-\frac{2(\beta+1)}{2\beta+1}+\delta}\ell^{\frac{\beta+3}{2\beta+1}}\Big\} + O_p\Big\{n^{-1}b^{-\frac{1}{4(\beta+1)}-\delta}\ell^{\frac{2\beta+5}{4(\beta+1)}+5\delta}\Big\}\\
&= -(b\ell)^{-1/2}xf(\tilde{\xi}_n) + o_p\Big\{n^{-\frac{\beta-3}{\beta-1}+\delta} + n^{-\frac{3\beta-1}{4(\beta+1)}+\delta} + n^{-\frac{\beta(2\beta-3)}{(\beta-1)(2\beta+1)}+\delta}\ell^{\frac{2(\beta+3)}{(\beta-1)(2\beta+1)}}\\
&\quad + n^{-\frac{4\beta+5}{4(\beta+1)}+\delta}\ell^{\frac{\beta+3}{2(\beta+1)}} + n^{-\frac{2(\beta+1)}{2\beta+1}+\delta}\ell^{\frac{\beta+3}{2\beta+1}} + n^{-\frac{4\beta^2+3\beta+1}{2(2\beta+1)(\beta+1)}+\delta}\ell^{\frac{\beta+3}{2(2\beta+1)}}\Big\}\\
&\quad + O_p\Big\{(b\ell)^{-1} + n^{-1}b^{-\frac{1}{4(\beta+1)}-\delta}\ell^{\frac{2\beta+5}{4(\beta+1)}+5\delta} + n^{-\frac{\beta-1}{2(\beta+1)}+\delta}(b\ell)^{-(1+\delta)/4}\Big\}.
\end{aligned}\tag{A7}$$

Note that (A7) holds under exponential mixing for any arbitrarily large β. Applying Lemma 4, we have, for arbitrarily small $\delta > 0$, that

$$\begin{aligned}
&\mathrm{pr}\Big[F_n^*\big\{\tilde{\xi}_n + (b\ell)^{-1/2}x\big\} \le p\Big|X_1,\ldots,X_n\Big]\\
&= \mathrm{pr}\Big((b\ell)^{1/2}\big[F_n^*\big\{\tilde{\xi}_n + (b\ell)^{-1/2}x\big\} - \tilde{F}_n\big\{\tilde{\xi}_n + (b\ell)^{-1/2}x\big\}\big]\\
&\qquad\qquad \le (b\ell)^{1/2}\big[p - \tilde{F}_n\big\{\tilde{\xi}_n + (b\ell)^{-1/2}x\big\}\big]\Big|X_1,\ldots,X_n\Big)\\
&= \Phi\Big((b\ell)^{1/2}\big[p - \tilde{F}_n\big\{\tilde{\xi}_n + (b\ell)^{-1/2}x\big\}\big]/\sigma\big\{\tilde{\xi}_n + (b\ell)^{-1/2}x\big\}\Big)\\
&\quad + \begin{cases} O_p\big\{\ell^{-1} + \ell^{1/2}n^{-1/2} + (b\ell)^{-1/2}\ell^{\delta}\big\} & \text{if } \ell = O(b) \text{ and } \alpha(t) = O(t^{-\beta}),\\ O_p\big\{\ell^{-1} + \ell^{1/2}n^{-1/2} + (b\ell)^{-1/2}\big\} & \text{if } \alpha(t) = O(e^{-Ct}).\end{cases}
\end{aligned}\tag{A8}$$

It follows from (A7), (A8) and Lemma 3(i) that for arbitrarily small $\delta > 0$,

$$\begin{aligned}
&\mathrm{pr}\Big[F_n^*\{\tilde{\xi}_n + (b\ell)^{-1/2}x\} \le p \Big| X_1, \ldots, X_n\Big] \\
&= \Phi\{-xf(\xi_p)/\sigma(\xi_p)\} + O_p\Big\{\ell^{-1} + \ell^{1/2}n^{-1/2} + (b\ell)^{-1/2}\ell^{\delta} + n^{-\frac{\beta-1}{2(\beta+1)}+\delta}(b\ell)^{(1-\delta)/4} \\
&\quad + n^{-1}b^{\frac{2\beta+1}{4(\beta+1)}-\delta}\ell^{\frac{4\beta+7}{4(\beta+1)}+5\delta}\Big\} + o_p\Big\{n^{-\frac{\beta-3}{\beta-1}+\delta}(b\ell)^{1/2} + n^{-\frac{3\beta-1}{4(\beta+1)}+\delta}(b\ell)^{1/2} \\
&\quad + n^{-\frac{\beta(2\beta-3)}{(\beta-1)(2\beta+1)}+\delta}b^{\frac{1}{2}}\ell^{\frac{1}{2}+\frac{2(\beta+3)}{(\beta-1)(2\beta+1)}} + n^{-\frac{4\beta+5}{4(\beta+1)}+\delta}b^{\frac{1}{2}}\ell^{\frac{\beta+2}{\beta+1}} + n^{-\frac{2(\beta+1)}{2\beta+1}+\delta}b^{\frac{1}{2}}\ell^{\frac{4\beta+7}{2(2\beta+1)}} \\
&\quad + n^{-\frac{4\beta^2+3\beta+1}{2(2\beta+1)(\beta+1)}+\delta}b^{\frac{1}{2}}\ell^{\frac{3\beta+4}{2(2\beta+1)}}\Big\}
\end{aligned} \tag{A9}$$

if $\beta \in (5, \infty)$ and $\ell = O(b)$, and

$$\begin{aligned}
&\mathrm{pr}\Big[F_n^*\{\tilde{\xi}_n + (b\ell)^{-1/2}x\} \le p \Big| X_1, \ldots, X_n\Big] \\
&= \Phi\{-xf(\xi_p)/\sigma(\xi_p)\} + O_p\Big\{\ell^{-1} + \ell^{1/2}n^{-1/2} + (b\ell)^{-1/2} + n^{-1}b^{\frac{1}{2}-\delta}\ell^{1+5\delta} \\
&\quad + n^{-\frac{1}{2}+\delta}(b\ell)^{(1-\delta)/4}\Big\} + o_p\Big\{n^{-\frac{3}{4}+\delta}(b\ell)^{1/2} + n^{-1+\delta}b^{\frac{1}{2}}\ell\Big\}
\end{aligned} \tag{A10}$$

under exponential mixing. Theorem 1 then follows by (A1), (A9), (A10) and noting that

$$\begin{aligned}
\mathrm{pr}\Big[F_n^*\{\tilde{\xi}_n + (b\ell)^{-1/2}x\} > p \Big| X_1, \ldots, X_n\Big] &\le \mathrm{pr}\Big\{(b\ell)^{1/2}(\xi_n^* - \tilde{\xi}_n) \le x \Big| X_1, \ldots, X_n\Big\} \\
&\le \mathrm{pr}\Big[F_n^*\{\tilde{\xi}_n + (b\ell)^{-1/2}x\} \ge p \Big| X_1, \ldots, X_n\Big].
\end{aligned}$$

[*Received on* 2 *January* 2017. *Editorial decision on* 1 *April* 2017]

Biometrika (2018), **103**, 1, *pp*. 1–9
Printed in Great Britain Advance Access publication on 31 July 2018

Supplementary material for Block bootstrap optimality and empirical block selection for sample quantiles with dependent data

BY T. A. KUFFNER
Department of Mathematics and Statistics, Washington University in St. Louis St. Louis, Missouri 63130, U.S.A.
kuffner@wustl.edu

S. M. S. LEE
Department of Statistics and Actuarial Science, The University of Hong Kong Pokfulam Road, Hong Kong
smslee@hku.hk

AND G. A. YOUNG
Department of Mathematics, Imperial College London, London SW7 2AZ U.K.
alastair.young@imperial.ac.uk

1. SOME TECHNICAL REMARKS

The type of weak dependence we assume is strong mixing. Because this is not the only type of weak dependence in common use (Wu, 2005a, 2011; Dedecker & Prieur, 2005; Doukhan & Louhichi, 1999), some justification for the choice of strong mixing is warranted. First, the vast majority of relevant block bootstrap literature relies on strong mixing assumptions; see, e.g., the monographs on block bootstrap (Lahiri, 2003) and subsampling (Politis et al., 1999). As a second reason, there is a rich literature which exploits the mixing rate to differentiate asymptotic orders when studying higher-order asymptotic properties of the bootstrap for statistical functionals with dependent data (Götze & Hipp, 1983; Lahiri, 2007).

Sample quantiles are often associated with empirical process theory, and the Bahadur representation of sample quantiles for dependent data can facilitate the study of limit theory (Sen, 1972; Wu, 2005b). However, while the Bahadur representation and corresponding empirical process theory is often useful for proving basic results such as consistency, asymptotic normality, or the law of the iterated logarithm, the question of optimal convergence rates requires a different approach utilizing higher-order asymptotics. We are not aware of any established empirical process results which can lead to the asymptotic orders derived herein. Specifically, the order, as a function of the available sample size n, of the remainder term in the Bahadur representation of sample quantiles under strong mixing is not necessarily informative about the asymptotic order of error in the block bootstrap approximation. Even if the remainder can be made as small as $n^{-3/4}$, this error will be transmitted to an error of order $n^{1/2}n^{-3/4} = n^{-1/4}$ when approximating the cumulative distribution function of the sample quantile by that of the empirical distribution. Such an error term will swamp the more precise orders established using our approach: see Figure 1 in § 4 of the paper, which shows that the optimal order is far below $n^{-1/4}$ for most values of the mixing rate parameter β. The main difference in our approach is that we study the event

$\{n^{1/2}(\text{sample quantile} - \text{population quantile}) \leq x\}$ directly without having to approximate the quantile process by the empirical process via the Bahadur representation. The latter approximation causes an unnecessary loss of precision in assessing the convergence rate of the distribution.

2. BACKGROUND

To understand how our results fit into the bootstrap landscape, consider the following. A broad categorization of settings for bootstrap methods is suggested by Lahiri (2003): (i) smooth functionals of independent data; (ii) nonsmooth functionals of independent data; (iii) smooth functionals of dependent data; and (iv) nonsmooth functionals of dependent data. Setting (i) is the classic setting of the bootstrap, with Hall (1992) being an authoritative reference. In setting (ii), bootstrap methods for approximating distributions of sample quantiles have been studied by Efron (1979); Bickel & Freedman (1981); Singh (1981); Babu (1986); Efron (1982); Ghosh et al. (1984); Hall & Sheather (1988); Hall et al. (1989); Hall & Martin (1991); De Angelis et al. (1993) and Falk & Janas (1992).

In setting (iii), i.e. smooth functionals of dependent data, the existing literature is concentrated on block bootstrap methods, beginning with Hall (1985), and Carlstein (1986). Subsequently, the moving block bootstrap was proposed by Künsch (1989) and Liu & Singh (1992). Other variants of the block bootstrap for smooth functionals have been suggested by Paparoditis & Politis (2001); Politis & Romano (1992, 1994) and Politis et al. (1997), to name a few. The various existing block bootstrap methods and their properties for weakly dependent sequences have been investigated by, for example, Bühlmann (1994); Naik-Nimbalkar & Rajarshi (1994); Hall et al. (1995); Götze & Künsch (1996); Lahiri (1992, 1996, 1999) and Bühlmann & Künsch (1999).

A distinct literature exists concerning the Bahadur representation of sample quantiles for strong mixing sequences. When such a representation exists, i.e. there is a nice relationship between the sample quantiles and empirical distribution function, then block bootstrap consistency properties will be implied by consistency properties of bootstrapping the empirical (quantile) process. Some recent work in this area is due to Sharipov & Wendler (2013), who proved weak consistency of circular block bootstrap under some additional conditions regarding mixing, block length and differentiability.

Another recent development in dependent data bootstrap methodology is the convolved subsampling bootstrap (Tewes et al., 2017). This bootstrap estimator is defined by the k-fold self-convolution of a subsampling distribution. In the special case of the sample means problem, this corresponds to our hybrid bootstrap. For the sample quantile problem which is the particular focus here, convolved subsampling bootstrap essentially computes the average of within-block sample quantiles over the b resampled blocks. By contrast, our estimator is the sample quantile of a single series formed by joining b blocks. Further theoretical comparison of these approaches will be undertaken elsewhere.

Other indirectly-related work includes Lahiri (2005), who studied consistency of jackknife-after-bootstrap variance estimation for bootstrap quantiles, and Gregory et al. (2015), who showed that the Sun & Lahiri (2006) strong consistency results for distribution and variance estimation via the moving block bootstrap also hold for the smoothed extended tapered block bootstrap (SETBB). The SETBB has been further developed for quantile regression by Gregory et al. (2018), though again only consistency is established, rather than optimal convergence rates. We mention that Shao & Politis (2013) employed a fixed b subsampling procedure to estimate confidence sets for statistics adhering to the smooth function model. In Section 3 they also consider functionals of the form $T(F)$ that can be linearly approximated as an average of influence functions. The quantile $T(F) = F^{-1}(p)$ would be covered by this framework, and their simulation study reported includes the median as an example. However, all of the papers cited above do not go beyond first-order consistency.

3. A FURTHER EXAMPLE

In this example, we construct a process whose mixing coefficients decay at a polynomial rate, but not an exponential rate. This is accomplished through Theorem 2.1 of Chanda (1974); see also Bandyopadhyay (2006) and §3 of Chen et al. (2016).

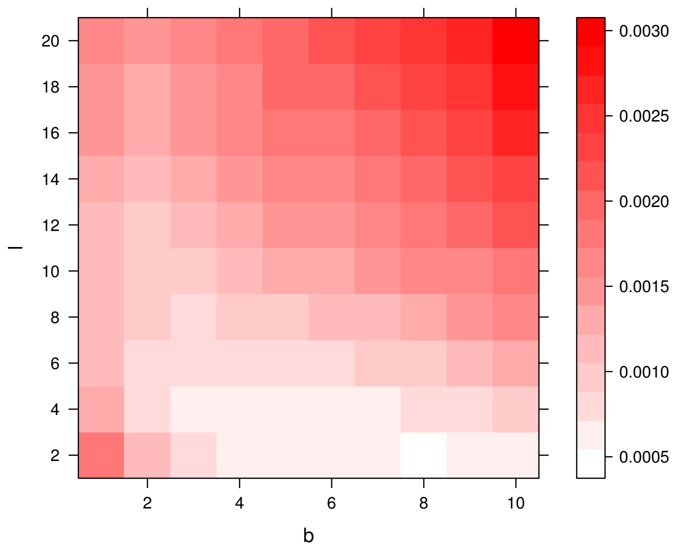
Example 3 (*Polynomial Mixing Rate*). Let the sequence $\{X_t\}_{t\in\mathbb{Z}}$ be generated according to

$$X_t = \sum_{j=0}^{\infty} c_j Z_{t-j},$$

where the Z_i are independent, identically distributed $N(0,1)$ and $c_j = (\frac{1}{j+1})^{\nu}$. Then X_t is strong mixing with a polynomial rate, and Chanda (1974) may be used to deduce that $\beta < \nu - 2$.

In practice, we cannot simulate from the above process exactly because it is expressed as an infinite series. Therefore, we approximate the process by truncating the series at 100 terms, which means that in reality X_t is approximated by a very high order MA process. For our numerical example, $\nu = 10.0$ and $n = 200$. As before, we consider $p = 1/2$, corresponding to $\xi_p = 0$. Simulation yields an approximation to the true value $G_n(2) \approx 0.95229$. The heatmap for this example shown in Figure 1 is based on 10,000 replications, and 10,000 bootstrap samples for each replication of the experiment. As with the previous two examples, the heatmap supports our

Fig. 1. Heatmap for the polynomial mixing rate example; $n = 200$.



finding of suboptimality of the choices of (b,ℓ) suggested by the subsampling bootstrap and the MBB.

4. PROOFS

Proof of Lemma 2:

Define $m_n = \left\lceil n^{\frac{\beta-1}{2(\beta+1)}-3\delta}\epsilon^{1-\frac{1}{2(\beta+1)}-\delta}\right\rceil$ and $\epsilon_n = \epsilon/m_n$. Then we have

$$\sup_{x\in\mathscr{B}_\epsilon(\xi_p)}\left|F_n(x)-F(x)\right| \le \max_{k\in\{0,\pm1,\ldots,\pm m_n\}}\left|F_n(\xi_p+k\epsilon_n)-F(\xi_p+k\epsilon_n)\right| + C\epsilon_n. \tag{1}$$

For each $k\in\{0,\pm1,\ldots,\pm m_n\}$, application of Lemma 1 with $V_{n,t}=\mathbf{1}\{X_t\le\xi_p+k\epsilon_n\}$, $p_n = F(\xi_p+k\epsilon_n)$ and $\delta_n = \left(\epsilon n^{2-\Delta_1}\right)^{\frac{-1}{1+\beta(1-\Delta_2)}}$, for arbitrarily small $\Delta_1,\Delta_2>0$, yields

$$\begin{aligned}&\mathbb{P}\left\{\left|F_n(\xi_p+k\epsilon_n)-F(\xi_p+k\epsilon_n)\right|>\epsilon_n\right\}\\ &\le Cn^{\frac{4}{2+\beta}}\exp\left\{-Cn^{(3-c_0)\delta}\right\} + C\epsilon^{-\frac{1+3\beta+2\beta^2}{2(1+\beta)^2}-\frac{\delta}{\beta+1}-C\Delta_2}n^{-\frac{\beta-1}{2(\beta+1)}-\frac{(\beta+3)\delta}{\beta+1}+C\Delta_2}.\end{aligned}$$

It follows by Bonferroni's inequality that

$$\begin{aligned}&\mathbb{P}\left\{\max_{k\in\{0,\pm1,\ldots,\pm m_n\}}\left|F_n(\xi_p+k\epsilon_n)-F(\xi_p+k\epsilon_n)\right|>\epsilon_n\right\}\\ &\le (2m_n+1)\mathbb{P}\left\{\left|F_n(\xi_p+k\epsilon_n)-F(\xi_p+k\epsilon_n)\right|>\epsilon_n\right\}\\ &\le Cn^{\frac{4}{2+\beta}+\frac{\beta-1}{2(\beta+1)}-3\delta}\exp\left\{-Cn^{(3-c_0)\delta}\right\}+Cn^{-\left\{\frac{(4-c_0)\beta+6-2c_0}{\beta+1}\right\}\delta+C\Delta_2} = o(1)\end{aligned}$$

for sufficiently small Δ_2, uniformly over $\epsilon\in\left[n^{-c_0},1\right)$. This, in conjunction with (1), implies that $\sup_{x\in\mathscr{B}_\epsilon(\xi_p)}\left|F_n(x)-F(x)\right| = O_p(\epsilon_n)$, which proves part (i) of Lemma 2.

To prove part (ii), write $n'=n-\ell+1$ and note that

$$\begin{aligned}F_n(x)-\tilde{F}_n(x) &= \frac{1}{\ell n'}\sum_{i=1}^{\ell-1}\left(\frac{\ell n'}{n}-i\right)\left\{\mathbf{1}\{X_i\le x\}+\mathbf{1}\{X_{n+1-i}\le x\}-2F(x)\right\}\\ &\qquad -\frac{\ell-1}{nn'}\sum_{i=\ell}^{n'}\left\{\mathbf{1}\{X_i\le x\}-F(x)\right\}.\end{aligned}\tag{2}$$

Define, for $c_0>1/2$, $m=n^{-\frac{c_0}{4(1+\beta)}-\frac{c_0\delta}{2}}\ell^{\frac{3\beta+5}{4(\beta+1)}+\frac{3\delta}{2}}$. It is clear that $m=o(\ell)$. Noting that for sufficiently large n and sufficiently small $\Delta>0$,

$$\frac{\ln\ell}{\ln n} > \frac{4\beta+7}{6(3\beta+5)}+\Delta,$$

we have

$$\begin{aligned}\frac{\ln m}{\ln n} &> -c_0\left\{\frac{1}{4(\beta+1)}+\frac{\delta}{2}\right\}+\left\{\frac{3\beta+5}{4(\beta+1)}+\frac{3\delta}{2}\right\}\left\{\frac{4\beta+7}{6(3\beta+5)}+\Delta\right\}\\ &= -(c_0-1/2)\left\{\frac{1}{4(\beta+1)}+\frac{\delta}{2}\right\}+\frac{1}{6}+\frac{(3\beta+5)\Delta}{4(\beta+1)}+\frac{\delta}{4}\left\{\frac{\beta+2}{3\beta+5}+6\Delta\right\}.\end{aligned}$$

We may therefore choose c_0 sufficiently close to $1/2$ and some $K < 6$ such that $\ln n / \ln m \le K$ and $c_0 K < 3$. It follows that for any $\epsilon \ge n^{-c_0}$,

$$\epsilon \ge n^{-c_0} \ge m^{-c_0 K} > \ell^{-c_0 K}.$$

Consider

$$\begin{aligned}
&\frac{1}{\ell n'} \sum_{i=1}^{\ell-1} \left\{ \frac{\ell n'}{n} - i \right\} \big(\mathbf{1}\{X_i \le x\} - F(x) \big) \\
&= \frac{1}{n} \sum_{i=1}^{\ell-1} \big\{ \mathbf{1}\{X_i \le x\} - F(x) \big\} - \frac{1}{\ell n'} \sum_{j=1}^{\ell-m} \left[\sum_{i=j}^{\ell-1} \big\{ \mathbf{1}\{X_i \le x\} - F(x) \big\} \right] \\
&\quad - \frac{1}{\ell n'} \sum_{j=\ell-m+1}^{\ell-1} \left[\sum_{i=j}^{\ell-1} \big\{ \mathbf{1}\{X_i \le x\} - F(x) \big\} \right] = I_1 - I_2 - I_3, \text{ say.}
\end{aligned}$$

Applying part (i), we have, uniformly over $\epsilon \in \big[n^{-c_0}, 1 \big)$, that

$$I_1 = O_p \left\{ n^{-1} \ell^{\frac{\beta+3}{2(\beta+1)} + 3\delta} \epsilon^{\frac{1}{2(\beta+1)} + \delta} \right\}$$

and

$$I_2 = O_p \left\{ \frac{1}{\ell n'} \sum_{j=1}^{\ell-m} (\ell - j)^{\frac{\beta+3}{2(\beta+1)} + 3\delta} \epsilon^{\frac{1}{2(\beta+1)} + \delta} \right\} = O_p \left\{ n^{-1} \ell^{\frac{\beta+3}{2(\beta+1)} + 3\delta} \epsilon^{\frac{1}{2(\beta+1)} + \delta} \right\}.$$

It is clear that

$$\begin{aligned}
I_3 &= O_p \big\{ (n\ell)^{-1} m^2 \big\} \\
&= O_p \left\{ n^{-1 - \frac{c_0}{2(1+\beta)} - c_0 \delta} \ell^{\frac{\beta+3}{2(\beta+1)} + 3\delta} \right\} \\
&= O_p \left\{ n^{-1} \ell^{\frac{\beta+3}{2(\beta+1)} + 3\delta} \epsilon^{\frac{1}{2(\beta+1)} + \delta} \right\}.
\end{aligned}$$

The bounds on I_1, I_2, I_3 therefore imply that

$$\begin{aligned}
\frac{1}{\ell n'} \sum_{i=1}^{\ell-1} \left(\frac{\ell n'}{n} - i \right) \big\{ \mathbf{1}\{X_i \le x\} + \mathbf{1}\{X_{n+1-i} \le x\} - 2F(x) \big\} \\
= O_p \left\{ n^{-1} \ell^{\frac{\beta+3}{2(\beta+1)} + 3\delta} \epsilon^{\frac{1}{2(\beta+1)} + \delta} \right\}.
\end{aligned} \tag{3}$$

It follows by part (i) again that

$$\begin{aligned}
\frac{\ell - 1}{n n'} \sum_{i=\ell}^{n'} \big\{ \mathbf{1}\{X_i \le x\} - F(x) \big\} &= O_p \left\{ \ell n^{-1 - \frac{\beta-1}{2(\beta+1)} + 3\delta} \epsilon^{\frac{1}{2(\beta+1)} + \delta} \right\} \\
&= o_p \left\{ n^{-1} \ell^{\frac{\beta+3}{2(\beta+1)} + 3\delta} \epsilon^{\frac{1}{2(\beta+1)} + \delta} \right\}.
\end{aligned} \tag{4}$$

Part (ii) then follows by combining (2), (3) and (4).

For the proof of part (iii), define $M_n = \left\lceil n^{\frac{\beta-1}{2(\beta+1)}-\delta}\epsilon^{(1-\delta)/2}\right\rceil$ and $\epsilon_n = \epsilon/M_n$. Then we have

$$\begin{aligned}
&\sup_{x\in\mathscr{B}_\epsilon(\xi_p)}\big|F_n(x)-F_n(\xi_p)-F(x)+p\big| \\
&\leq \max_{k\in\{0,\pm1,\dots,\pm M_n\}}\big|F_n(\xi_p+k\epsilon_n)-F_n(\xi_p)-F(\xi_p+k\epsilon_n)+p\big|+C\epsilon_n. \qquad (5)
\end{aligned}$$

For each $k\in\{\pm1,\dots,\pm m_n\}$, set, for the application of Lemma 1, $V_{n,t} = \big|\mathbf{1}\{X_t\leq\xi_p+k\epsilon_n\}-\mathbf{1}\{X_t\leq\xi_p\}\big|$, $p_n = \big|F(\xi_p+k\epsilon_n)-p\big|$ and $\delta_n = \big(n^{2-\Delta_1}\epsilon^{\Delta_2}\big)^{-1/(\beta(1-\Delta_2)+1)}$, for arbitrarily small $\Delta_1,\Delta_2>0$. Noting that $C^{-1}\epsilon_n\leq p_n\leq C\epsilon$, we have, by Lemma 1, that for any $\Delta\in(0,1)$,

$$\begin{aligned}
&\mathbb{P}\left\{\Big|F_n(\xi_p+k\epsilon_n)-F_n(\xi_p)-F(\xi_p+k\epsilon_n)+p\Big|>\epsilon_n\right\} \\
&\leq C\big(n^{2-\Delta_1}\epsilon^{\Delta_2}\big)^{1/(\beta(1-\Delta_2)+1)}\exp\left\{-C\big(n^{2-\Delta_1}\epsilon^{\Delta_2}\big)^{-1/(\beta(1-\Delta_2)+1)}n^{\frac{2}{\beta+1}+2\delta}\epsilon^{\delta}\right\} \\
&\quad +Cn^{\frac{3\beta+1}{2(\beta+1)}-\delta}\epsilon^{-(1+\delta)/2+\Delta}\big(n^{2-\Delta_1}\epsilon^{\Delta_2}\big)^{-\beta(1-\Delta)/(\beta(1-\Delta_2)+1)}.
\end{aligned}$$

It follows by Bonferroni's inequality that for sufficiently small $\Delta'>0$ and for any $c_0\in(0,2)$,

$$\begin{aligned}
&\mathbb{P}\left\{\max_{k\in\{0,\pm1,\dots,\pm m_n\}}\big|F_n(\xi_p+k\epsilon_n)-F_n(\xi_0)-F(\xi_p+k\epsilon_n)+p\big|>\epsilon_n\right\} \\
&\leq(2M_n+1)\mathbb{P}\left\{\Big|F_n(\xi_p+k\epsilon_n)-F_n(\xi_0)-F(\xi_p+k\epsilon_n)+p\Big|>\epsilon_n\right\} \\
&\leq Cn^{\frac{\beta+3}{2(\beta+1)}}\exp\left\{-Cn^{(2-c_0)\delta/2}\right\}+Cn^{-(2-c_0)\delta+\Delta'}=o(1),
\end{aligned}$$

uniformly over $\epsilon\in\big[n^{-c_0},1\big)$. This, in conjunction with (5), implies that $\sup_{x\in\mathscr{B}_\epsilon(\xi_p)}\big|F_n(x)-F_n(\xi_p)-F(x)+p\big| = O_p(\epsilon_n)$, which proves part (iii) of Lemma 2.

Proof of Lemma 3:
Let $c_0>1/2$ be as specified in Lemma 2(ii). Define, for $\epsilon\in[n^{-c_0},1)$,

$$\delta_n(\epsilon) = n^{-1}\epsilon^{\frac{1}{2(\beta+1)}+\delta}\ell^{\frac{\beta+3}{2(\beta+1)}+3\delta},$$

$$\epsilon_1 = \left\{n^{-1}\ell^{\frac{\beta+3}{2(\beta+1)}+3\delta}\right\}^{\frac{2(\beta+1)}{2\beta+1-2\delta(\beta+1)}} \quad\text{and}\quad \epsilon_2 = n^{-1/2}+\epsilon_1.$$

Note that $\delta_n(\epsilon_1)=\epsilon_1$. Using Lemma 2(ii), we have, for some $\alpha\in(0,1)$, any $M>0$ and sufficiently large $n,\tilde{M}$,

$$\begin{aligned}
&\mathbb{P}\big(\tilde{\xi}_n-\xi_p>M\epsilon_2\big) \\
&\leq\mathbb{P}\Big\{\sup_{x\in\mathscr{B}_{M\epsilon_2}(\xi_p)\cap\mathscr{N}_p}|\tilde{F}_n(x)-F_n(x)|>\tilde{M}\delta_n(M\epsilon_2)\Big\} \\
&\qquad+\mathbb{P}\big\{F_n(\xi_p+M\epsilon_2)\leq p+\tilde{M}\delta_n(M\epsilon_2)\big\} \\
&\leq\tilde{M}^{-1}+\mathbb{P}\Big(n^{1/2}\big\{F_n(\xi_p+M\epsilon_2)-F(\xi_p+M\epsilon_2)\big\} \\
&\qquad\leq-2^{-1}Mf(\xi_p)\big[1+n^{1/2}\epsilon_1-CM^{-(1-\alpha)}\tilde{M}n^{1/2}\big\{\delta_n(n^{-1/2})+\epsilon_1\big\}\big]\Big). \qquad (6)
\end{aligned}$$

Note that if $\epsilon_1 = o(n^{-1/2})$, then $\delta_n(n^{-1/2}) = o(n^{-1/2})$, so that

$$\lim_{n\to\infty} -2^{-1}Mf(\xi_p)\big[\big]1 + n^{1/2}\epsilon_1 - CM^{-(1-\alpha)}\tilde{M}n^{1/2}\big\{\delta_n(n^{-1/2}) + \epsilon_1\big\}\big] = -2^{-1}Mf(\xi_p).$$

If $\epsilon_1 \geq C_0 n^{-1/2}$ for some $C_0 > 0$, then $\delta_n(n^{-1/2}) \leq C_0^{-\alpha}\delta_n(\epsilon_1) = C_0^{-\alpha}\epsilon_1$, so that for M sufficiently large, we have

$$\begin{aligned} &n^{1/2}\epsilon_1 - CM^{-(1-\alpha)}\tilde{M}n^{1/2}\big\{\delta_n(n^{-1/2}) + \epsilon_1\big\} \\ &\geq C_0\Big\{1 - CM^{-(1-\alpha)}\tilde{M}\big(C_0^{-\alpha} + 1\big)\Big\} \geq C_0/2. \end{aligned}$$

The above results suggest that

$$-2^{-1}Mf(\xi_p)\big[1 + n^{1/2}\epsilon_1 - CM^{-(1-\alpha)}\tilde{M}n^{1/2}\big\{\delta_n(n^{-1/2}) + \epsilon_1\big\}\big] \leq -4^{-1}Mf(\xi_p)$$

for M, n sufficiently large. It then follows from (6) that $\mathbb{P}\big(\tilde{\xi}_n - \xi_p > M\epsilon_2\big)$ can be made arbitrarily small if we choose n, M and $\tilde{M}$ large enough, using the fact that $F_n(\xi_p + M\epsilon_2) - F(\xi_p + M\epsilon_2) = O_p(n^{-1/2})$. The same arguments can be applied to the lower tail probability $\mathbb{P}\big(\tilde{\xi}_n - \xi_p < -M\epsilon_2\big)$. Thus we have $\tilde{\xi}_n = \xi_p + O_p(\epsilon_2)$. Lemma 3(i) then follows as δ can be made arbitrarily small.

To prove part (ii), write

$$\begin{aligned} \varepsilon_0(n,\ell) &= n^{-1/2} + n^{-\frac{2(\beta+1)}{2\beta+1}}\ell^{\frac{\beta+3}{2\beta+1}}, \\ \varepsilon_1(n,\ell) &= n^{-\frac{4\beta+5}{4(\beta+1)}}\ell^{\frac{\beta+3}{2(\beta+1)}} + n^{-\frac{2(\beta+1)}{2\beta+1}}\ell^{\frac{\beta+3}{2\beta+1}}, \\ \varepsilon_2(n,\ell) &= n^{-\frac{\beta-3}{\beta-1}} + n^{-\frac{\beta(2\beta-3)}{(\beta-1)(2\beta+1)}}\ell^{\frac{2(\beta+3)}{(\beta-1)(2\beta+1)}}, \\ \varepsilon(n,\ell) &= \varepsilon_1(n,\ell) + \varepsilon_2(n,\ell). \end{aligned}$$

Denote by $X_{(1)} \leq \cdots \leq X_{(n)}$ the ordered sequence of $X_1, \ldots, X_n$. For any arbitrarily small $\Delta, \Delta' > 0$, $\mathbb{P}\big\{\tilde{F}_n(\tilde{\xi}_n) \geq p + n^\delta\varepsilon(n,\ell)\big\}$ is bounded above by

$$\begin{aligned} &\mathbb{P}\Big\{F_n(X_{(j+1)}) = \cdots = F_n(X_{(j+k)}) \geq p + n^\delta\varepsilon(n,\ell) - n^\Delta\varepsilon_1(n,\ell), \\ &\qquad F_n(X_{(j)}) < p + n^\Delta\varepsilon_1(n,\ell),\ \big|X_{(j+1)} - \xi_p\big| \leq n^{\Delta'}\varepsilon_0(n,\ell) \text{ some } j,k \geq 1\Big\} \qquad (7) \\ &\qquad + \mathbb{P}\Big\{\big|\tilde{\xi}_n - \xi_p\big| > n^{\Delta'}\varepsilon_0(n,\ell)\Big\} + \mathbb{P}\Big\{\sup\big|\tilde{F}_n(x) - F_n(x)\big| > n^\Delta\varepsilon_1(n,\ell)\Big\}, \end{aligned}$$

where the supremum in the last probability is taken over $x \in \mathscr{B}_{n^{\Delta'}\varepsilon_0(n,\ell)}(\xi_p) \cap \mathscr{N}_p$. Noting that $n^{4-\beta}\varepsilon_0(n,\ell)^2 = O\big\{\varepsilon_2(n,\ell)^{\beta-1}\big\} = O\big\{\varepsilon(n,\ell)^{\beta-1}\big\}$, we have

$$\begin{aligned} &n^{-(\beta-2-\Delta')}\varepsilon_0(n,\ell)\big\{n^\delta\varepsilon(n,\ell)\big\}^{-(\beta-1)} \\ &= O\Big\{n^{-(\beta-2-\Delta')-\delta(\beta-1)}\varepsilon_0(n,\ell)n^{\beta-4}\varepsilon_0(n,\ell)^{-2}\Big\} \\ &= O\Big\{n^{-2+\Delta'-\delta(\beta-1)}\varepsilon_0(n,\ell)^{-1}\Big\} \\ &= o\Big\{n^{-2-\Delta'}\varepsilon_0(n,\ell)^{-1}\Big\} \end{aligned}$$

for sufficiently small $\Delta' > 0$. Thus we may find a positive sequence $\{\eta_n\}$ satisfying

$$\eta_n = o\big\{n^{-2-\Delta'}\varepsilon_0(n,\ell)^{-1}\big\} \text{ and } n^{-(\beta-2-\Delta')}\varepsilon_0(n,\ell)\big\{n^\delta\varepsilon(n,\ell)\big\}^{-(\beta-1)} = o(\eta_n) \qquad (8)$$

for sufficiently small $\Delta' > 0$. Noting that $n^{\Delta}\varepsilon_1(n,\ell) = o\big(n^{\delta}\varepsilon(n,\ell)\big)$ for sufficiently small Δ, following the proof of Lemma 5.4(iv) of Sun and Lahiri (2006) and using (8), the first probability in (7) can be bounded above, for sufficiently large n and sufficiently small $\Delta' > 0$, by

$$\begin{aligned} & n \sum_{j>2^{-1}n^{1+\delta}\varepsilon(n,\ell)} \mathbb{P}\Big\{X_1 = X_j \in \mathscr{B}_{n^{\Delta'}\varepsilon_0(n,\ell)}(\xi_p) \cap \mathscr{N}_p\Big\} \\ & \quad \le Cn^{2+\Delta'}\varepsilon_0(n,\ell)\eta_n + Cn^{2-\beta+\Delta'}\varepsilon_0(n,\ell)\big\{n^{\delta}\varepsilon(n,\ell)\big\}^{-(\beta-1)}\eta_n^{-1} = o(1). \end{aligned}$$

That the last two probabilities in (7) converge to 0 follows from Lemma 3(i) and Lemma 2(ii), respectively. From the above results we derive that $\mathbb{P}\big\{\tilde{F}_n(\tilde{\xi_n}) \ge p + n^{\delta}\varepsilon(n,\ell)\big\} = o(1)$. Similar arguments show also that $\mathbb{P}\big\{\tilde{F}_n(\tilde{\xi_n}) \le p - n^{\delta}\varepsilon(n,\ell)\big\} = o(1)$, which completes the proof of part (ii).

REFERENCES

BABU, G. (1986). A note on bootstrapping the variance of sample quantile. *Annals of the Institute of Statistical Mathematics* **38**, 439–443.

BANDYOPADHYAY, S. (2006). A note on strong mixing. *Technical report* .

BICKEL, P. & FREEDMAN, D. (1981). Some asymptotic theory for the bootstrap. *Annals of Statistics* **9**, 1196–1217.

BÜHLMANN, P. (1994). Blockwise bootstrapped empirical process for stationary sequences. *Annals of Statistics* **22**, 995–1012.

BÜHLMANN, P. & KÜNSCH, H. R. (1999). Block length selection in the bootstrap for time series. *Computational Statistics and Data Analysis* **31**, 295–310.

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

CHANDA, K. (1974). Strong mixing properties of linear stochastic processes. *Journal of Applied Probability* **11**, 401–408.

CHEN, X., SHAO, Q.-M., WU, W. B. & XU, L. (2016). Self-normalized Cramér-type moderate deviations under dependence. *Annals of Statistics* **44**, 1593–1617.

DE ANGELIS, D., HALL, P. & YOUNG, G. A. (1993). A note on coverage error of bootstrap confidence intervals for quantiles. *Mathematical Proceedings of the Cambridge Philosophical Society* **114**, 517–531.

DEDECKER, J. & PRIEUR, C. (2005). New dependence coefficients. Examples and applications to statistics. *Probab. Theory Related Fields* **132**, 203–236.

DOUKHAN, P. & LOUHICHI, S. (1999). A new weak dependence condition and applications to moment inequalities. *Stochastic Process. Appl.* **84**, 313–342.

EFRON, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics* **7**, 1–26.

EFRON, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*. Society for Industrial and Applied Mathematics, Philadelphia.

FALK, M. & JANAS, J. (1992). Edgeworth expansions for studentized and prepivoted sample quantiles. *Statistics and Probability Letters* **14**, 13–24.

GHOSH, M., PARR, W., SINGH, K. & BABU, G. (1984). A note on bootstrapping the sample median. *Annals of Statistics* **12**, 1130–1135.

GÖTZE, F. & HIPP, C. (1983). Asymptotic expansions for sums of weakly dependent random vectors. *Z. Wahrsch. Verw. Gebiete* **64**, 211–239.

GÖTZE, F. & KÜNSCH, H. R. (1996). Second-order correctness of the blockwise bootstrap for stationary observations. *Annals of Statistics* **24**, 1914–1933.

GREGORY, K. B., LAHIRI, S. N. & NORDMAN, D. J. (2015). A smooth block bootstrap for statistical functionals and time series. *Journal of Time Series Analysis* **36**, 442–461.

GREGORY, K. B., LAHIRI, S. N. & NORDMAN, D. J. (2018). A smooth block bootstrap for quantile regression with time series. *Annals of Statistics* **46**, 1138–1166.

HALL, P. (1985). Resampling a coverage pattern. *Stochastic Processes and their Applications* **20**, 231–246.

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

HALL, P., DICICCIO, T. J. & ROMANO, J. P. (1989). On smoothing and the bootstrap. *Annals of Statistics* **17**, 692–704.

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

HALL, P. & MARTIN, M. A. (1991). On error incurred using the bootstrap variance estimate when constructing confidence intervals for quantiles. *Journal of Multivariate Analysis* **38**, 70–81.

HALL, P. & SHEATHER, S. (1988). On the distribution of a studentized quantile. *Journal of the Royal Statistical Society Series B* **50**, 381–391.

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

LAHIRI, S. N. (1992). Edgeworth correction by moving block bootstrap for stationary and nonstationary data. In *Exploring the Limits of Bootstrap*, R. LePage & L. Billard, eds. Wiley, pp. 263–270.

LAHIRI, S. N. (1996). On Edgeworth expansion and moving block bootstrap for studentized m-estimators in multiple linear regression models. *Journal of Multivariate Analysis* **56**, 42–59.

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

LAHIRI, S. N. (2003). *Resampling Methods for Dependent Data*. Springer-Verlag.

LAHIRI, S. N. (2005). Consistency of the jackknife-after-bootstrap variance estimator for the bootstrap quantiles of a studentized statistic. *Annals of Statistics* **33**, 2475–2506.

LAHIRI, S. N. (2007). Asymptotic expansions for sums of block-variables under weak dependence. *Ann. Statist.* **35**, 1324–1350.

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

NAIK-NIMBALKAR, U. & RAJARSHI, M. (1994). Validity of blockwise bootstrap for empirical processes with stationary observations. *Annals of Statistics* **22**, 980–994.

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

POLITIS, D. & ROMANO, J. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *Annals of Statistics* **22**, 2031–2050.

POLITIS, D., ROMANO, J. & WOLF, M. (1997). Subsampling for heteroskedastic time series. *Journal of Econometrics* **81**, 281–317.

POLITIS, D. & ROMANO, J. P. (1992). A circular block resampling procedure for stationary data. In *Exploring the Limits of Bootstrap*, R. LePage & L. Billard, eds. Wiley, pp. 263–270.

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

SEN, P. K. (1972). On Bahadur representation of sample quantile for sequences of ϕ-mixing random variables. *Journal of Multivariate Analysis* **2**, 77–95.

SHAO, X. & POLITIS, D. N. (2013). Fixed b subsampling and the block bootstrap: improved confidence sets based on p-value calibration. *Journal of the Royal Statistical Society Series B* **75**, 161–184.

SHARIPOV, O. S. & WENDLER, M. (2013). Normal limits, nonnormal limits, and the bootstrap for quantiles of dependent data. *Statistics and Probability Letters* **83**, 1028–1035.

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

SUN, S. & LAHIRI, S. N. (2006). Bootstrapping the sample quantile of a weakly dependent sequence. *Sankhyā* **68**, 130–166.

TEWES, J., POLITIS, D. N. & NORDMAN, D. J. (2017). Convolved subsampling estimation with applications to block bootstrap. *arXiv:1706.07237* .

WU, W. B. (2005a). Nonlinear system theory: another look at dependence. *Proc. Natl. Acad. Sci. USA* **102**, 14150–14154.

WU, W. B. (2005b). On the Bahadur representation of sample quantiles for dependent sequences. *Ann. Statist.* **33**, 1934–1963.

WU, W. B. (2011). Asymptotic theory for stationary processes. *Stat. Interface* **4**, 207–226.

[*Received on* 2 *January* 2017. *Editorial decision on* 1 *April* 2017]