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 (MBB), 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 (MBB). The hybrid block bootstrap is seen to achieve remarkable improvements in accuracy for sample quantile distribution estimation compared to the subsampling bootstrap and MBB.

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 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 MBB, 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 MBB 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. 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 MBB (Künsch, 1989). Therefore, the consistency results in Kuffner et al. (2018) are for a hybrid between the MBB 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 (MSE) 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 i.i.d. 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 the 'hybrid' bootstrap distribution estimator for sample quantiles. It is a hybrid between the MBB ($b = \lfloor n/\ell \rfloor$) 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 MBB ($b = \lfloor n/\ell \rfloor$). 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

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 $\{X_i\}_{i\in\mathbb{Z}}$.

3.1. *The Block Bootstrap*

The moving blocks bootstrap (MBB) (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. 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 $\geq h$. 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 MBB 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 MBB, 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 MBB 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)

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

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, i.e. 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 $x\in\mathbb{R}$, $\sigma(x)^2=\lim_{n\to\infty}\mathrm{Var}\big(n^{1/2}F_n(x)\big)=\sum_{t=-\infty}^{\infty}\mathrm{Cov}\,\big(\mathbf{1}\{X_0\le x\},\mathbf{1}\{X_t\le x\}\big)$. 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 $\{1, \ldots, n-\ell+1\}$,

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

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

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

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

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

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

We may deduce from Theorem 1 the following:

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 $\delta > 0$,

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

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 MBB 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.

Case (ii). Exponential mixing.

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

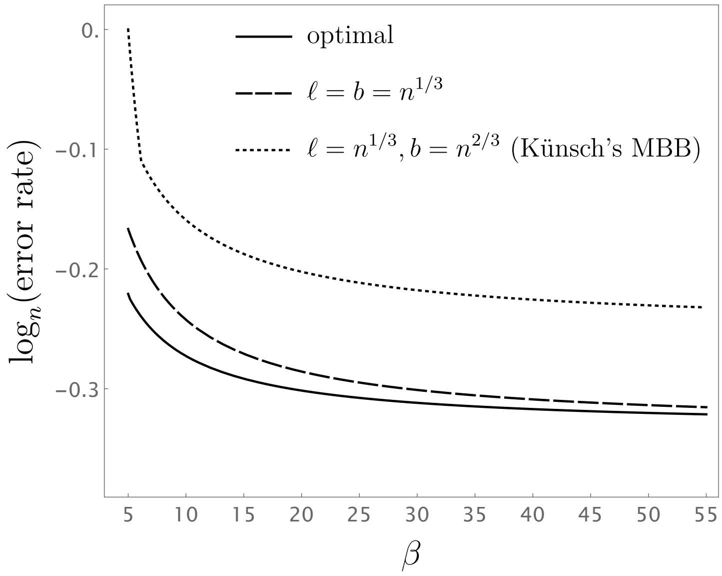
$$\mathbb{P}\Big((b\ell)^{1/2}\big(\xi_n^* - \tilde{\xi}_n\big) \leq x \Big| X_1, \ldots, X_n\Big) = \mathbb{P}\Big(n^{1/2}\big(\hat{\xi}_n - \xi_p\big) \leq x\Big) + O_p\Big(n^{-1/3+\delta}\Big),$$

for any arbitrarily small $\delta > 0$. For MBB, 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 circular 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, which gives the convergence rate of the block bootstrap distribution estimator for $n^{1/2}\big(F_n(x) - F(x)\big)$ and has a bearing on the regular smooth function model. Consider the simpler case of exponential mixing. It is easily seen

Fig. 1. From Theorem 1, Case (i). 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 MBB choice $(b = n^{2/3}, \ell = n^{1/3})$ is optimal for Künsch's MBB.



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 MBB 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 $n^{1/2}\big(\hat{\xi}_n - \xi_p\big)$, embodied in Theorem 1, differ substantially from the regular case in that local estimation of F over a shrinking neighbourhood of size $O_p\big((b\ell)^{-1/2}\big)$ around ξ_p incurs an error of order $n^{-1/2}(b\ell)^{1/4}$, which favours a small b and precludes MBB 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 (8) in § 8. 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 Section 8 for more details.

5. RELEVANCE TO COVERAGE ERROR

Define

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

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 (8) (Lahiri & Sun, 2009) from § 8, we obtain that

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

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 (8) 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 $\rho \ge 1$,

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

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

$$\hat{G}_M^{(j)}(x) - G_M(x) = \Delta^{(j)}\big(M, \lfloor c_1 M^{1/3} \rfloor, \lfloor c_2 M^{1/3} \rfloor\big) + O(M^{-1/2}), \quad (7)$$

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 (8), (5) and (7), we have

$$\begin{aligned} \hat{G}_M^{(j)}(x) - \hat{G}_n(x) &= G_M(x) + \Delta^{(j)}(M, \lfloor c_1 M^{1/3} \rfloor, \lfloor c_2 M^{1/3} \rfloor) + O(M^{-1/2}) \\ &\qquad - G_n(x) - \Delta(n, \lfloor c_1 n^{1/3} \rfloor, \lfloor c_2 n^{1/3} \rfloor) - O(n^{-1/2}) \\ &= \Delta^{(j)}(M, \lfloor c_1 M^{1/3} \rfloor, \lfloor c_2 M^{1/3} \rfloor) + O(M^{-1/2}). \end{aligned}$$

It follows that

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

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. For concreteness, we consider $p = 1/2$, and simulate the mean squared errors (MSEs) 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 heatmaps 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 MBB for the sample sizes we consider. This facilitates comparison with the MBB choice of $b = \lfloor n/\ell \rfloor$ for a range of values of ℓ. In particular, we give values for ℓ approximately equal to $n^{1/2}$ (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 MBB choice* $b=\lfloor n/\ell\rfloor$.

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

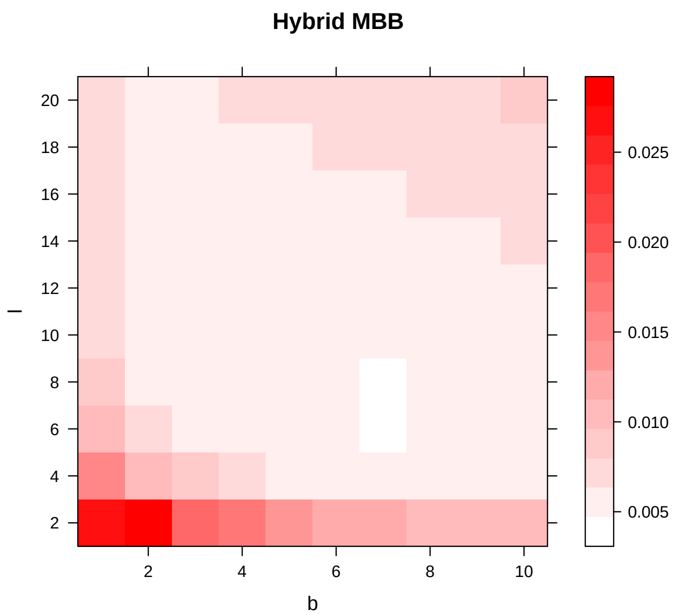
Example 1 (*ARMA(1,1)*). Suppose that the observations are generated according to an ARMA (1,1) model

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

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 MSE 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 MSE 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 MSE is 0.00468, with $(b,\ell)=(7,8)$. By contrast, the minimum MSE for the MBB is 0.00637, with $(b,\ell)=(33,6)$, and the subsampling bootstrap, which fixes $b=1$, has minimum MSE of 0.00754, with $\ell=14$.

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



We also compute the values of the pair (b,ℓ) which minimize MSE for other sample sizes, $n=500,1000$, and 2000. These results are shown in Table 2. Comparing with Table 1, we note

Table 2. *ARMA(1,1) model. Choices of* (b, ℓ) *which minimise the MSE 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

that the MSE-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 MSE-minimizing value of b is much larger than 1.

The theory says that the hybrid MBB has an error rate in estimation of $G_n(1)$ of $n^{-1/3}$, so we should expect the MSE to decrease at rate $n^{-2/3}$. In fact, a regression of $\log(MSE)$ on $\log(n)$ for the values reported in Table 2 has slope -0.6885, which is not far off $-2/3$. The heatmap illustrates that the subsampling and MBB 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 MBB if, instead, we are interested in estimation of $\mathbb{P}\{n^{1/2}\big(F_n(x) - F(x)\big) \leq y\}$. This was verified by considering, for all combinations of (b, ℓ), the MSE of the estimator $\mathbb{P}\{(b\ell)^{1/2}\big(F_n^*(x) - \tilde{F}_n(x)\big) \leq y \Big| X_1, \ldots, X_n\}$, 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 MSE achieved by MBB is 0.00084, with $(b, \ell) = (25, 4)$. This is very similar to the overall minimum MSE of 0.00082, seen for $(b, \ell) = (18, 5)$. The minimum MSE 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 MSE of MBB is then 0.00108, with $(b, \ell) = (28, 7)$, with the same minimum MSE for the hybrid block bootstrap, achieved for $(b, \ell) = (30, 6)$. Here the subsampling bootstrap yields an optimal MSE of 0.00227 when $\ell = 8$. These illustrative figures confirm that the hybrid block bootstrap has little advantage over MBB 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,

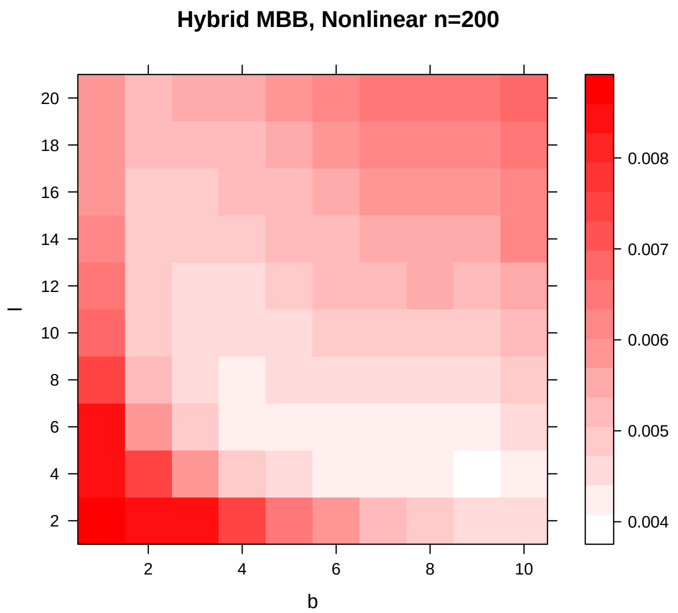
$$Y_t = X_t^2.$$

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

$$\mathbb{P}(Y_t = X_t^2 \le \xi_p) = 1/2,$$

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

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



The heatmap of Figure 3 shows again that the subsampling and MBB choices of (b, ℓ) are suboptimal from the perspective of minimizing MSE.

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 MSE in estimation of $G_n(1)$ over 2500 replications for each combination of (c_1, c_2). By contrast, the MSE 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 an MSE that is 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 MSE of the adaptively chosen estimator over the 2500 replications was observed

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

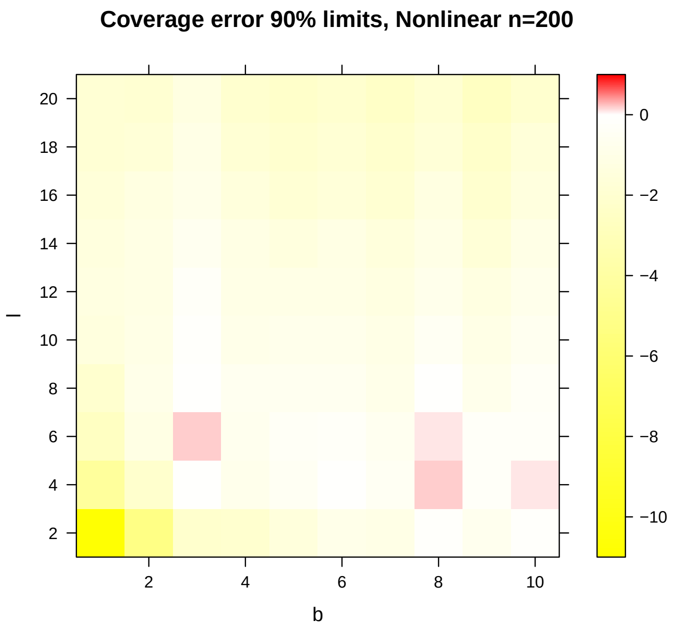


Table 3. *Nonlinear (squared) ARMA(2,3) model: MSE in estimation of $G_n(1)$ over 2500 replications, for $b = \lfloor c_1 n^{1/3} \rfloor$ and $\ell = \lfloor c_2 n^{1/3} \rfloor$, $n = 512$. MSE of adaptive procedure was 0.00189.*

		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

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.

In 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 MBB.

8. 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 $x \in \mathbb{R}$,

$$\mathbb{P}\Big(n^{1/2}\big(\hat{\xi}_n - \xi_p\big) \leq x\Big) = \Phi\big(xf(\xi_p)/\sigma(\xi_p)\big) + O(n^{-1/2}). \tag{8}$$

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

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

LEMMA 1. *Let $\{V_{n,t} : t = 0, \pm 1, \pm 2, \dots\}$ be a double array of row-wise stationary strong mixing Bernoulli (p_n) random variables with $0 < p_n \leq q < 1$ 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 $\epsilon_n = o(1)$, $n^{-1} \leq \delta_n = o(1)$ and any $\delta \in (0,1)$, we have*

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

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

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

(iii) $\displaystyle\sup_{x \in \mathscr{B}_\epsilon(\xi_p) \cap \mathscr{N}_p} \big|F_n(x) - F_n(\xi_p) - F(x) + p\big| = O_p\left(n^{-\frac{\beta-1}{2(\beta+1)} + \delta} \epsilon^{(1+\delta)/2} \right)$ *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 $\mathscr{K} \subset \mathbb{R}$,*

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

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

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

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

$$\begin{aligned} \hat{\kappa}_j(x) &= (b\ell)^{j/2}b^{1-j}g_j(\mathcal{V}_1, \ldots, \mathcal{V}_j) \\ &= (b\ell)^{j/2}b^{1-j}\big\{g_j(\mathbb{E}\mathcal{V}_1, \ldots, \mathbb{E}\mathcal{V}_j) + O_p\big(n^{-1/2}\ell^{(1-\beta)/2} + n^{-1/2}\ell^{1-j}|\mathcal{A}_{2j}|^{1/2}\big)\big\}, \quad (9) \end{aligned}$$

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 $\delta > 0$,

$$g_j(\mathbb{E}\mathcal{V}_1, \ldots, \mathbb{E}\mathcal{V}_j) = O\big(\ell^{-\beta} + \ell^{1-j+\delta}\big). \quad (10)$$

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

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

it can be shown by induction and (10) that

$$\mathcal{A}_j = O\big(\ell^{j-1-\beta} + \ell^{\delta} + \ell^{(j-2)/2-(1/2-\delta)\mathbf{1}\{j \text{ odd}\}}\big), \quad j \ge 3. \quad (11)$$

It follows from (9), (10) and (11) that

$$\begin{aligned} \hat{\kappa}_2(x) &= \ell g_2(\mathcal{V}_1, \mathcal{V}_2) = \ell(\mathcal{V}_2 - \mathcal{V}_1^2) \\ &= \sum_{1\le|t|\le\ell-1}(1 - |t|/\ell)\,\mathrm{Cov}\big(\mathbf{1}\{X_0 \le x\}, \mathbf{1}\{X_t \le x\}\big) + O_p\big(n^{-1/2}\{\ell^{1+(1-\beta)/2} + |\mathcal{A}_4|^{1/2}\}\big) \\ &= \sigma(x)^2 + O(\ell^{-1}) + O_p\big(n^{-1/2}\ell^{(3-\beta)/2} + n^{-1/2}\ell^{1/2}\big) \\ &= \sigma(x)^2 + O_p\big(\ell^{-1} + n^{-1/2}\ell^{1/2}\big) \quad (12) \end{aligned}$$

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

$$\begin{aligned} \hat{\kappa}_j(x) &= (b\ell)^{j/2}b^{1-j}\big\{g_j(\mathbb{E}\mathcal{V}_1, \ldots, \mathbb{E}\mathcal{V}_j) + O_p\big(n^{-1/2}\ell^{(1-\beta)/2} + n^{-1/2}\ell^{1-j}|\mathcal{A}_{2j}|^{1/2}\big)\big\} \\ &= b^{-(j-2)/2} \times O_p\big(\ell^{j/2-\beta} + \ell^{1-j/2+\delta} + n^{-1/2}\ell^{(j+1-\beta)/2} + n^{-1/2}\ell^{1/2}\big) \\ &= O_p\big(b^{-1/2}\ell^{-1/2+\delta} + n^{-1/2}\ell^{(3-\beta)/2} + n^{-1/2}b^{-1/2}\ell^{1/2}\big). \quad (13) \end{aligned}$$

Without imposing the condition $\ell = O(b)$, the above arguments can similarly be applied to the case of exponential mixing rates to establish (12) and a stronger version of (13), 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 $(b\ell)^{1/2}\{F_n^*(x) - \tilde{F}_n(x)\}$ and $\Phi\big(\,\cdot\,/\sigma(x)\big)$ yields, for any arbitrarily large $C' > 0$,

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

where $\hat{\kappa}_x^*(t)$ denotes the conditional characteristic function of $(b\ell)^{1/2}\{F_n^*(x) - \tilde{F}_n(x)\}/\sigma(x)$. Lemma 4 then follows by bounding $\hat{\kappa}_x^*(t) + t^2/2$ using (12) and (13) under polynomial mixing, or using (12) and the stronger version of (13) 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 $\tilde{\xi}_n$,

$$\begin{aligned}
&p - \tilde{F}_n\big(\tilde{\xi}_n + (b\ell)^{-1/2}x\big) \\
&= \Big\{p - \tilde{F}_n(\tilde{\xi}_n)\Big\} + \Big\{\tilde{F}_n(\tilde{\xi}_n) - \tilde{F}_n\big(\tilde{\xi}_n + (b\ell)^{-1/2}x\big)\Big\} \\
&= F_n(\tilde{\xi}_n) - F_n\big(\tilde{\xi}_n + (b\ell)^{-1/2}x\big) + o_p\Big(n^{-\frac{\beta-3}{\beta-1}+\delta} + n^{-\frac{\beta(2\beta-3)}{(\beta-1)(2\beta+1)}+\delta}\ell^{\frac{2(\beta+3)}{(\beta-1)(2\beta+1)}} \\
&\qquad + n^{-\frac{4\beta+5}{4(\beta+1)}+\delta}\ell^{\frac{\beta+3}{2(\beta+1)}} + n^{-\frac{2(\beta+1)}{2\beta+1}+\delta}\ell^{\frac{\beta+3}{2\beta+1}}\Big) + O_p\Big(n^{-1}b^{-\frac{1}{4(\beta+1)}-\delta}\ell^{\frac{2\beta+5}{4(\beta+1)}+5\delta}\Big) \\
&= -(b\ell)^{-1/2}xf(\tilde{\xi}_n) + o_p\Big(n^{-\frac{\beta-3}{\beta-1}+\delta} + n^{-\frac{3\beta-1}{4(\beta+1)}+\delta} + n^{-\frac{\beta(2\beta-3)}{(\beta-1)(2\beta+1)}+\delta}\ell^{\frac{2(\beta+3)}{(\beta-1)(2\beta+1)}} \\
&\qquad + n^{-\frac{4\beta+5}{4(\beta+1)}+\delta}\ell^{\frac{\beta+3}{2(\beta+1)}} + n^{-\frac{2(\beta+1)}{2\beta+1}+\delta}\ell^{\frac{\beta+3}{2\beta+1}} + n^{-\frac{4\beta^2+3\beta+1}{2(2\beta+1)(\beta+1)}+\delta}\ell^{\frac{\beta+3}{2(2\beta+1)}}\Big) \\
&\qquad + O_p\left((b\ell)^{-1} + n^{-1}b^{-\frac{1}{4(\beta+1)}-\delta}\ell^{\frac{2\beta+5}{4(\beta+1)}+5\delta} + n^{-\frac{\beta-1}{2(\beta+1)}+\delta}(b\ell)^{-(1+\delta)/4}\right).
\end{aligned} \tag{14}$$

Note that (14) holds under exponential mixing for any arbitrarily large β. Applying Lemma 4, we have, for arbitrarily small $\delta > 0$, that

$$\begin{aligned}
&\mathbb{P}\Big(F_n^*\big(\tilde{\xi}_n + (b\ell)^{-1/2}x\big) \le p \Big| X_1,\ldots,X_n\Big) \\
&= \mathbb{P}\Big((b\ell)^{1/2}\{F_n^*\big(\tilde{\xi}_n + (b\ell)^{-1/2}x\big) - \tilde{F}_n\big(\tilde{\xi}_n + (b\ell)^{-1/2}x\big)\} \\
&\qquad\quad \le (b\ell)^{1/2}\{p - \tilde{F}_n\big(\tilde{\xi}_n + (b\ell)^{-1/2}x\big)\}\Big| X_1,\ldots,X_n\Big) \\
&= \Phi\Big((b\ell)^{1/2}\{p - \tilde{F}_n\big(\tilde{\xi}_n + (b\ell)^{-1/2}x\big)\}/\sigma\big(\tilde{\xi}_n + (b\ell)^{-1/2}x\big)\Big) \\
&\qquad + \begin{cases} O_p\big(\ell^{-1} + \ell^{1/2}n^{-1/2} + (b\ell)^{-1/2}\ell^{\delta}\big) & \text{if } \ell = O(b) \text{ and } \alpha(t) = O(t^{-\beta}), \\ O_p\big(\ell^{-1} + \ell^{1/2}n^{-1/2} + (b\ell)^{-1/2}\big) & \text{if } \alpha(t) = O(e^{-Ct}). \end{cases}
\end{aligned} \tag{15}$$

It follows from (14), (15) and Lemma 3(i) that for arbitrarily small $\delta > 0$,

$$
\begin{aligned}
&\mathbb{P}\Big(F_n^*\big(\tilde{\xi}_n + (b\ell)^{-1/2}x\big) \le p \Big| X_1, \ldots, X_n\Big) \\
&= \Phi\big(-xf(\xi_p)/\sigma(\xi_p)\big) + O_p\Big(\ell^{-1} + \ell^{1/2}n^{-1/2} + (b\ell)^{-1/2}\ell^{\delta} + n^{-\frac{\beta-1}{2(\beta+1)}+\delta}(b\ell)^{(1-\delta)/4} \\
&\quad + n^{-1}b^{\frac{2\beta+1}{4(\beta+1)}-\delta}\ell^{\frac{4\beta+7}{4(\beta+1)}+5\delta}\Big) + o_p\Big(n^{-\frac{\beta-3}{\beta-1}+\delta}(b\ell)^{1/2} + n^{-\frac{3\beta-1}{4(\beta+1)}+\delta}(b\ell)^{1/2} \\
&\quad + n^{-\frac{\beta(2\beta-3)}{(\beta-1)(2\beta+1)}+\delta}b^{\frac{1}{2}}\ell^{\frac{1}{2}+\frac{2(\beta+3)}{(\beta-1)(2\beta+1)}} + n^{-\frac{4\beta+5}{4(\beta+1)}+\delta}b^{\frac{1}{2}}\ell^{\frac{\beta+2}{\beta+1}} + n^{-\frac{2(\beta+1)}{2\beta+1}+\delta}b^{\frac{1}{2}}\ell^{\frac{4\beta+7}{2(2\beta+1)}} \\
&\quad + n^{-\frac{4\beta^2+3\beta+1}{2(2\beta+1)(\beta+1)}+\delta}b^{\frac{1}{2}}\ell^{\frac{3\beta+4}{2(2\beta+1)}}\Big)
\end{aligned} \tag{16}
$$

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

$$
\begin{aligned}
&\mathbb{P}\Big(F_n^*\big(\tilde{\xi}_n + (b\ell)^{-1/2}x\big) \le p \Big| X_1, \ldots, X_n\Big) \\
&= \Phi\big(-xf(\xi_p)/\sigma(\xi_p)\big) + O_p\Big(\ell^{-1} + \ell^{1/2}n^{-1/2} + (b\ell)^{-1/2} + n^{-1}b^{\frac{1}{2}-\delta}\ell^{1+5\delta} \\
&\quad + n^{-\frac{1}{2}+\delta}(b\ell)^{(1-\delta)/4}\Big) + o_p\Big(n^{-\frac{3}{4}+\delta}(b\ell)^{1/2} + n^{-1+\delta}b^{\frac{1}{2}}\ell\Big)
\end{aligned} \tag{17}
$$

under exponential mixing. Theorem 1 then follows by (8), (16), (17) and noting that

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

9. Discussion

In the absence of exact, finite-sample results, accurate estimation of quantiles is essential for implementation of statistical inference procedures. As sample quantiles are nonsmooth functionals, conventional bootstrap theory for the smooth function model does not apply to estimation of their distribution. In dependent data settings, with some notable exceptions, little is known about the block bootstrap for distribution estimation of sample quantiles. In this paper we have established a general optimality theory for block bootstrap procedures in such settings under strong mixing conditions, and we have shown that a hybrid block bootstrap is optimal, in the sense of having the fastest convergence rate for distribution estimation. In addition, of course, since the hybrid block bootstrap is based on bootstrap samples of smaller size than the data sample, it provides computational advantage over MBB. How one should choose (b, ℓ) in a given application to capture the good theoretical properties of the hybrid block bootstrap requires further consideration. We have provided discussion of an empirical scheme that seems fruitful for this purpose and which will be further developed and refined elsewhere. Future work will also study the SETBB 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 the SETBB method, though this latter procedure is complicated by additional tuning parameters.

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.

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