Journal of the
Royal Statistical Society
SERIES B Statistical Methodology

J. R. Statist. Soc. B (2013)
75, *Part* 1, *pp.*

Fixed *b* subsampling and the block bootstrap: improved confidence sets based on *p*-value calibration

Xiaofeng Shao

University of Illinois at Urbana–Champaign, Champaign, USA

and Dimitris N. Politis

University of California at San Diego, La Jolla, USA

[Received June 2011. Revised April 2012]

Summary. Subsampling and block-based bootstrap methods have been used in a wide range of inference problems for time series. To accommodate the dependence, these resampling methods involve a bandwidth parameter, such as the subsampling window width and block size in the block-based bootstrap. In empirical work, using different bandwidth parameters could lead to different inference results, but traditional first-order asymptotic theory does not capture the choice of the bandwidth. We propose to adopt the fixed *b* approach, as advocated by Kiefer and Vogelsang in the heteroscedasticity–auto-correlation robust testing context, to account for the influence of the bandwidth on inference. Under the fixed *b* asymptotic framework, we derive the asymptotic null distribution of the *p*-values for subsampling and the moving block bootstrap, and further propose a calibration of the traditional small-*b*-based confidence intervals (regions or bands) and tests. Our treatment is fairly general as it includes both finite dimensional parameters and infinite dimensional parameters, such as the marginal distribution function. Simulation results show that the fixed *b* approach is more accurate than the traditional small *b* approach in terms of approximating the finite sample distribution, and that the calibrated confidence sets tend to have smaller coverage errors than the uncalibrated counterparts.

Keywords: Block bootstrap; Calibration; Iterative bootstrap; Prepivoting; Subsampling

1. Introduction

Subsampling and block-based bootstrap methods have been widely used in inference problems for time series; see Politis *et al.* (1999a) and Lahiri (2003) for book length treatments of these important resampling methods. To accommodate the unknown time series dependence nonparametrically, these methods introduce a bandwidth parameter l_n, such as the block size in the block-based bootstrap and the subsampling window width in subsampling. The bandwidth l_n plays an important role in the finite sample performance of subsampling or block-bootstrap-based inference. Intuitively, if the bandwidth (or block size) is too small, it may not capture the dependence in a time series sufficiently, whereas, if it is too large, the number of blocks for subsampling or resampling is too small to lead to a good approximation of the finite sample distribution. Statistically speaking, the bandwidth l_n is a smoothing parameter as it usually leads to a bias–variance trade-off in variance estimation or size–power trade-off in testing on

Address for correspondence: Xiaofeng Shao, Department of Statistics, University of Illinois at Urbana–Champaign, Champaign, IL 61820, USA.
E-mail: xshao@uiuc.edu

1369–7412/13/75000

the basis of subsampling and the block bootstrap. In traditional asymptotic theory, $l_n \to \infty$ as the sample size $n \to \infty$ and the fraction $b = l_n/n \to 0$, which is a necessary condition for the general consistency of subsampling and block-based bootstrap methods without additional assumptions. Therefore, the role of l_n (or b) does not show up in the conventional first-order asymptotics, although in practice the choice of l_n does affect the subsampling or block bootstrap distribution estimator and related operating characteristics.

In this paper, we aim to offer a new perspective on the use of these smoothing-parameter-dependent resampling methods based on the so-called fixed b approach, which was first proposed by Kiefer and Vogelsang (2005) in the context of heteroscedasticity–auto-correlation robust testing. It was found that the asymptotic distribution that is obtained under the fixed b framework (i.e. $b \in (0, 1]$ is held fixed in the asymptotics) provides a better approximation to the sampling distribution of the Studentized test statistic than its counterpart obtained under the small b framework (i.e. $b \to 0$ as $n \to \infty$). See Jansson (2004) and Sun *et al.* (2008) for rigorous theoretical justifications. The fixed b approach has the advantage of accounting for the effect of the bandwidth, as different bandwidth parameters correspond to different limiting (null) distributions. The literature on inference using the fixed b approach and its variants has been growing steadily; see Hashimzade and Vogelsang (2008), Sun *et al.* (2008), Shao (2010a), Goncalves and Vogelsang (2011) and Sayginsoy and Vogelsang (2011) among others for recent contributions.

In this paper, we adopt Kiefer and Vogelsang's fixed b approach and investigate its possible gain in the context of subsampling (Politis and Romano, 1994) and the moving block bootstrap (MBB) (Künsch, 1989; Liu and Singh, 1992). The extension to other bandwidth-dependent bootstrap methods, such as the tapered block bootstrap (Paparoditis and Politis, 2001, 2002; Shao, 2010b) and the dependent wild bootstrap (Shao, 2010c) are possible but are not pursued here. Under the fixed b asymptotics, Lahiri (2001) showed that the subsampling and the MBB approximations are no longer consistent in the case of the sample mean, which seems to suggest that a direct application of the fixed b approach is fruitless. A novel feature of our extension is that we study the limiting null distribution of the p-value, which is $U(0, 1)$ (i.e. a uniform distribution on $[0, 1]$) under the small b asymptotics but is dependent on b and differs from $U(0, 1)$ under the fixed b asymptotics. For a scalar parameter, we calibrate the nominal coverage level on the basis of the pivotal limiting null distribution of the p-value under the fixed b framework, and modify the small-b-based confidence interval by inverting the corresponding test. Thus the effect of the bandwidth parameter l_n on the subsampling or block bootstrap distribution approximation is captured to the first order by using a p-value-based adjustment. Simulation studies are conducted to demonstrate that the fixed b approach delivers confidence intervals of better coverage in most situations and that the fixed-b-based intervals are slightly wider than the small b counterparts, which is consistent with early findings associated with the fixed b approach; see for example Kiefer and Vogelsang (2005).

So far the use of the fixed b approach has been restricted to the inference of a finite dimensional parameter. Since subsampling and the MBB have also been used in the inference of infinite dimensional parameters, such as the marginal distribution function of a stationary time series, we explore an extension of the fixed b idea to construct confidence bands for these infinite dimensional parameters. Unlike the case of a scalar parameter, the limiting null distribution of the subsampling-based p-value is not pivotal under the fixed b asymptotics and it depends on the unknown dependence structure of the underlying process. To alleviate the problem, we apply the subsampling method to approximate the sampling distribution of the p-value so that inference becomes feasible. This double-subsampling approach is also used in constructing the confidence region for a vector parameter.

The remainder of the paper is organized as follows. Section 2 introduces an extension of the fixed b approach to subsampling and the MBB in the mean case. Section 3 lays out a general framework and describes the fixed-b-based confidence interval (region) for a finite dimensional parameter and some implementational issues. Section 4 presents an extension of the fixed b approach to confidence band construction for the marginal distribution function. Simulation results are reported in Section 5. Section 6 concludes and Appendix A contains some technical details.

2. Inference for the mean

To help the reader to understand the essence of the fixed b approach, we shall focus on a simple problem: inference for the mean of a stationary time series. Suppose that we want to test $H_0:\mu=\mu_0$ *versus* $H_1:\mu\neq\mu_0$ on the basis of the observations $\{X_t\}_{t=1}^n$ from a univariate stationary time series with $\mathbb{E}(X_t)=\mu$. Under suitable moment and weak dependence conditions, we have $(\bar{X}_n-\mu)\sqrt{n}\rightarrow_{\mathrm{D}} N(0,\sigma^2)$ where $\sigma^2=\Sigma_{k\in\mathbb{Z}}\gamma(k)$ is the long-run variance with $\gamma(k)=\mathrm{cov}(X_0,X_k)$ and '$\rightarrow_{\mathrm{D}}$' denotes convergence in distribution. The scale parameter σ^2 can be consistently estimated by the so-called lag window estimator

$$\hat{\sigma}_n^2=\sum_{j=1-n}^{n-1} K(j/l)\hat{\gamma}_n(j),$$

where $l=l_n$ is a bandwidth parameter, $K(\cdot)$ is a kernel function and

$$\hat{\gamma}_n(j)=n^{-1}\sum_{k=|j|+1}^{n}(X_k-\bar{X}_n)(X_{k-|j|}-\bar{X}_n)$$

is the sample autocovariance at lag j.

A natural test statistic is $G_n=n(\bar{X}_n-\mu_0)^2/\hat{\sigma}_n^2$. To ensure the consistency of $\hat{\sigma}_n^2$ as an estimator of σ^2, the bandwidth parameter $l=bn$, where $b\in(0,1]$, typically satisfies $1/l+l/n=o(1)$ (i.e. $b+n^{-1}/b=o(1)$) as $n\rightarrow\infty$. These are the so-called small b asymptotics, under which the limiting null distribution of G_n is the distribution of χ_1^2. Under the fixed b asymptotics, the ratio of bandwidth to sample size b is held fixed and G_n converges in distribution (under the null hypothesis) to $U(b)$, whose detailed form can be found in Kiefer and Vogelsang (2005). The distribution of $U(b)$ depends on the kernel K and b, so different choices of the kernel and bandwidth lead to different limiting null distributions. From both empirical and theoretical perspectives, the fixed b approach has been shown to provide a more accurate approximation to the finite sample distribution of G_n than the small b counterpart under the null, so it corresponds to better size in hypothesis testing. Owing to the duality between confidence interval construction and hypothesis testing, the interval that is delivered by the fixed b approach tends to have an empirical coverage that is closer to the nominal coverage.

In the next two subsections, we describe an extension of the fixed b approach to subsampling and the MBB for the mean inference problem. A further extension to the inference of a finite dimensional parameter is made in Section 3. Throughout, we use $\lfloor a\rfloor$ to denote the integer part of $a\in\mathbb{R}$ and $\lceil a\rceil$ to denote the smallest integer larger than or equal to a. The symbol $N(\mu,\Sigma)$ denotes the normal distribution with mean μ and covariance matrix Σ.

2.1. Subsampling

For the inference of the mean, the subsampling method approximates the sampling distribution of $(\bar{X}_n-\mu)\sqrt{n}$ with the empirical distribution generated by its subsample counterpart

$(\bar{X}_{j,j+l-1}-\bar{X}_n)\surd l$, where $\bar{X}_{j,j+l-1}=l^{-1}\Sigma_{i=j}^{j+l-1}X_i$, $j=1,\ldots,N=n-l+1$. Let

$$L_{n,l}(x)=N^{-1}\sum_{j=1}^{N}\mathbf{1}\{(\bar{X}_{j,j+l-1}-\bar{X}_n)\surd l\leqslant x\}$$

be the subsampling approximation, where $\mathbf{1}(A)$ denotes the indicator function of the set A. For a given $\alpha\in(0,1)$ (say, $\alpha=0.05$ or $\alpha=0.1$), we define the subsampling-based critical values as $c_{n,l}(1-\alpha)=\inf\{x:L_{n,l}(x)\geqslant 1-\alpha\}$. Then a $100(1-\alpha)$% (one-sided) confidence interval is $[\bar{X}_n-n^{-1/2}c_{n,l}(1-\alpha),\infty)$ and the $100(1-\alpha)$% (two-sided) equal-tailed confidence interval is $[\bar{X}_n-n^{-1/2}c_{n,l}(1-\alpha/2),\bar{X}_n-n^{-1/2}c_{n,l}(\alpha/2)]$. In the testing context, if the alternative hypothesis is $H_1:\mu>\mu_0$, then we reject the null hypothesis at the significance level α if $\mu_0\notin[\bar{X}_n-n^{-1/2}c_{n,l}(1-\alpha),\infty)$; if the alternative hypothesis is $H_1:\mu<\mu_0$, then the null is rejected provided that $\mu_0\notin(-\infty,\bar{X}_n-n^{-1/2}c_{n,l}(\alpha)]$, which is also a one-sided confidence interval with nominal level $1-\alpha$. If the alternative hypothesis is $H_1:\mu\neq\mu_0$, then the null hypothesis is rejected when $\mu_0\notin[\bar{X}_n-n^{-1/2}c_{n,l}(1-\alpha/2),\bar{X}_n-n^{-1/2}c_{n,l}(\alpha/2)]$. The above inference is based on the traditional small-b-based asymptotic theory, under which $\sup_{x\in\mathbb{R}}|L_{n,l}(x)-\Phi(x/\sigma)|=o_p(1)$, where Φ is the distribution function for the standard normal distribution, and $\sup_{x\in\mathbb{R}}|L_{n,l}(x)-P\{(\bar{X}_n-\mu)\surd n\leqslant x\}|=o_p(1)$, i.e. the subsampling method provides a consistent approximation to the sampling distribution of $(\bar{X}_n-\mu)\surd n$ and its limiting distribution. Note that we are using the data-centred subsampling distribution for testing as recommended by Berg *et al.* (2010).

Under the fixed b framework, $L_{n,l}$ does not converge to $N(0,\sigma^2)$ in distribution. Instead, Lahiri (2001) showed that the limit of $L_{n,l}(x)$ is

$$(1-b)^{-1}\int_0^{1-b}\mathbf{1}[\{W(b+t)-W(t)-b\,W(1)\}\sigma/\surd b\in(-\infty,x]]\,\mathrm{d}t,$$

where $W(t)$ is standard Brownian motion. Since $L_{n,l}$ converges to a random measure, the subsampling-based inference is asymptotically invalid under the fixed b framework. To alleviate the problem, we shall modify the traditional subsampling-based inference procedure by considering the subsampling-based p-value and its limiting null distribution. For the one-sided alternative hypothesis $H_1:\mu>\mu_0$, we define the p-value as

$$\mathrm{pval}_{n,l}^{\mathrm{SUB}}:=\frac{1}{N}\sum_{j=1}^{N}\mathbf{1}\{(\bar{X}_n-\mu_0)\surd n\leqslant(\bar{X}_{j,j+l-1}-\bar{X}_n)\surd l\}.$$

Under the small b asymptotics, it can be shown that $\mathrm{pval}_{n,l}^{\mathrm{SUB}}$ converges to $U[0,1]$ in distribution under the null, whereas, under the fixed b asymptotics, its limiting null distribution is the distribution of $G(b)$, where

$$G(b)=(1-b)^{-1}\int_0^{1-b}\mathbf{1}[W(1)\leqslant\{W(b+t)-W(t)-b\,W(1)\}/\surd b]\,\mathrm{d}t.$$

Note that the nuisance parameter σ is cancelled out in $G(b)$, which is pivotal for a given b. Let $G_\alpha(b)$ denote the 100α% quantile of the distribution $G(b)$. Then, at the significance level α, we reject the null and favour the alternative hypothesis $H_1:\mu>\mu_0$, if the (realized) p-value is smaller than $G_\alpha(b)$. Correspondingly, a one-sided confidence interval under the fixed b asymptotics can be obtained by inverting the test, i.e.

$$\left\{\mu:\frac{1}{N}\sum_{j=1}^{N}\mathbf{1}\{(\bar{X}_n-\mu)\surd n\leqslant(\bar{X}_{j,j+l-1}-\bar{X}_n)\surd l\}\geqslant G_\alpha(b)\right\},$$

which is

$$\{\mu: (\bar{X}_n - \mu)\sqrt{n} \leqslant c_{n,l}\{1 - G_\alpha(b)\}\} = [\bar{X}_n - n^{-1/2} c_{n,l}\{1 - G_\alpha(b)\}, \infty).$$

Compared with the conventional subsampling-based confidence interval, the difference lies in the replacement of α by $G_\alpha(b)$ in $c_{n,l}$. Note that α is the $100\alpha\%$ quantile of $U(0,1)$, which is the limiting null distribution of the p-value under the small b asymptotics. In a similar manner, we can obtain the $100(1-\alpha)\%$ two-sided equal-tailed confidence interval for μ as $[\bar{X}_n - n^{-1/2} c_{n,l}\{1 - G_{\alpha/2}(b)\}, \bar{X}_n - n^{-1/2} c_{n,l}\{G_{\alpha/2}(b)\}]$ and another one-sided confidence interval $(-\infty, \bar{X}_n - n^{-1/2} c_{n,l}\{G_\alpha(b)\}]$ under the fixed b asymptotics. We can view the fixed-b-based inference as a way of calibrating the small b counterpart with the level α adjusted by $G_\alpha(b)$, so the effect of b on the inference is taken into account.

If we want to construct a symmetric two-sided confidence interval for μ, then we can approximate the sampling distribution of $|\bar{X}_n - \mu|\sqrt{n}$ by

$$\tilde{L}_{n,l}(x) = N^{-1} \sum_{j=1}^{N} \mathbf{1}(|\bar{X}_{j,j+l-1} - \bar{X}_n|\sqrt{l} \leqslant x).$$

Letting $\tilde{c}_{n,l}(1-\alpha) = \inf\{x : \tilde{L}_{n,l}(x) \geqslant 1-\alpha\}$, then the $100(1-\alpha)\%$ symmetric confidence interval for μ is $[\bar{X}_n - n^{-1/2} \tilde{c}_{n,l}(1-\alpha), \bar{X}_n + n^{-1/2} \tilde{c}_{n,l}(1-\alpha)]$ under the small b asymptotic theory. The p-value is defined as

$$\widetilde{\text{pval}}_{n,l}^{\text{SUB}} = \frac{1}{N} \sum_{j=1}^{N} \mathbf{1}\{|\bar{X}_n - \mu_0|\sqrt{n} \leqslant |\bar{X}_{j,j+l-1} - \bar{X}_n|\sqrt{l}\}.$$

Under the fixed b asymptotics, the limiting null distribution of $\widetilde{\text{pval}}_{n,l}^{\text{SUB}}$ is the distribution of $\tilde{G}(b)$, where

$$\tilde{G}(b) = (1-b)^{-1} \int_0^{1-b} \mathbf{1}\{|W(1)| \leqslant |W(b+t) - W(t) - b\,W(1)|/\sqrt{b}\}\,\mathrm{d}t.$$

Let $\tilde{G}_\alpha(b)$ denote the $100\alpha\%$ quantile of the distribution $\tilde{G}(b)$. Then the fixed-b-based $100(1-\alpha)\%$ symmetric confidence interval is

$$\{\mu : N^{-1} \sum_{j=1}^{N} \mathbf{1}(|\bar{X}_n - \mu|\sqrt{n} \leqslant |\bar{X}_{j,j+l-1} - \bar{X}_n|\sqrt{l}) \geqslant \tilde{G}_\alpha(b)\},$$

i.e.

$$[\bar{X}_n - n^{-1/2} \tilde{c}_{n,l}\{1 - \tilde{G}_\alpha(b)\}, \bar{X}_n + n^{-1/2} \tilde{c}_{n,l}\{1 - \tilde{G}_\alpha(b)\}]. \quad (1)$$

2.2. Moving block bootstrap

In this subsection, we shall consider the approximation of the sampling distribution of $(\bar{X}_n - \mu)\sqrt{n}$ by the MBB. Denote the MBB sample by $\{X_1^*(l), \ldots, X_n^*(l)\}$ with the dependence on the block size l being explicit. Then we approximate the sampling distribution of $(\bar{X}_n - \mu)\sqrt{n}$ by the conditional distribution of $\{\bar{X}_n^*(l) - \bar{X}_n\}\sqrt{n}$ or $[\bar{X}_n^*(l) - \mathbb{E}^*\{\bar{X}_n^*(l)\}]\sqrt{n}$ given the data, where $\bar{X}_n^*(l) = n^{-1} \sum_{t=1}^{n} X_t^*(l)$ is the sample mean for the bootstrap sample, and $\mathbb{E}^*$ and var* are used to denote the conditional expectation and variance respectively. To avoid the issue of centring, we could use the circular bootstrap (Politis and Romano, 1992), which is asymptotically equivalent to the MBB (Lahiri, 2003). For simplicity, we shall focus on the bootstrap approximation based on $\{\bar{X}_n^*(l) - \bar{X}_n\}\sqrt{n}$. The same idea can be applied to the other bootstrap approximation. We define the MBB-based p-value as

$$\text{pval}_{n,l}^{\text{MBB}} := E^*[\mathbf{1}[(\bar{X}_n - \mu_0)\sqrt{n} \leqslant \{\bar{X}_n^*(l) - \bar{X}_n\}\sqrt{n}]],$$

which corresponds to the alternative $H_1 : \mu > \mu_0$. Under the small b asymptotics, the p-value $\text{pval}_{n,l}^{\text{MBB}}$ is expected to converge to $U(0,1)$ in distribution under the null, although we are unaware of a formal proof. Under the fixed b asymptotics, we assume that $R_b = n/l = 1/b$ (i.e. the reciprocal of b) is an integer to ease our discussion. Then $\bar{X}_n^*(l) = n^{-1}\sum_{j=1}^{R_b l} X_j^*(l) = n^{-1}\sum_{j=1}^{R_b} v_j^*$, where, conditional on the data, $\{v_j^*\}_{j=1}^{R_b}$ are independent and identically distributed (IID) with a discrete uniform distribution, $P(v_1^* = \sum_{t=j}^{j+l-1} X_t) = 1/N$, $j = 1, \ldots, N$. Hence the above p-value is equal to

$$\frac{1}{N^{R_b}} \sum_{j_1, j_2, \ldots, j_{R_b}=1}^{N} \mathbf{1}\left\{ n^{-1/2} \sum_{h=1}^{R_b} \sum_{s=j_h}^{j_h+l-1} (X_s - \mu_0) - n^{1/2}(\bar{X}_n - \mu_0) \geqslant n^{1/2}(\bar{X}_n - \mu_0) \right\}.$$

Under the fixed b asymptotics and the null hypothesis, it converges in distribution to

$$H(b) := (1-b)^{-R_b} \int_0^{1-b} \cdots \int_0^{1-b} \mathbf{1}\left[\sum_{h=1}^{R_b} \{W(t_h + b) - W(t_h)\} \geqslant 2\,W(1) \right] \mathrm{d}t_1 \ldots \mathrm{d}t_{R_b}.$$

Let $H_\alpha(b)$ denote the $100\alpha\%$ quantile of $H(b)$. In practice, we usually further approximate the distribution of $\{\bar{X}_n^*(l) - \bar{X}_n\}\sqrt{n}$ by taking a finite number of bootstrap samples, say $X_t^{*,j}(l)\}_{t=1}^n$, $j = 1, \ldots, B$. We approximate the sampling distribution of $(\bar{X}_n - \mu)\sqrt{n}$ by

$$M_{n,l,B}^*(x) = \frac{1}{B} \sum_{j=1}^{B} \mathbf{1}[\{\bar{X}_n^{*,j}(l) - \bar{X}_n\}\sqrt{n} \leqslant x],$$

where $\bar{X}_n^{*,j}(l) = n^{-1}\sum_{t=1}^n X_t^{*,j}(l)$. Let $c_{n,l,B}^*(1-\alpha) = \inf\{x : M_{n,l,B}^*(x) \geqslant 1-\alpha\}$. The corresponding fixed-b-based two-sided equal-tailed confidence interval for μ is then

$$[\bar{X}_n - n^{-1/2} c_{n,l,B}^*\{1 - H_{\alpha/2}(b)\}, \bar{X}_n - n^{-1/2} c_{n,l,B}^*\{H_{\alpha/2}(b)\}]$$

and the one-sided confidence intervals can be formed analogously to those developed for the subsampling method. The details are omitted. The above discussion is based on the assumption that $R_b = 1/b$ is an integer. When R_b is not an integer, we use a fraction of the last resampled block to make the bootstrap sample size equal to original sample size. Then the p-value is

$$\frac{1}{N^{\lfloor R_b \rfloor}} \frac{1}{l\lfloor R_b \rfloor + 1} \sum_{j_1, j_2, \ldots, j_{\lfloor R_b \rfloor}=1}^{N} \sum_{j_{\lfloor R_b \rfloor+1}=1}^{l\lfloor R_b \rfloor+1} \mathbf{1}\left[n^{-1/2} \left\{ \sum_{h=1}^{\lfloor R_b \rfloor} \sum_{s=j_h}^{j_h+l-1} (X_s - \bar{X}_n) + \sum_{s=j_{\lfloor R_b \rfloor+1}}^{j_{\lfloor R_b \rfloor+1}+n-l\lfloor R_b \rfloor-1} (X_s - \bar{X}_n) \right\} \geqslant n^{1/2}(\bar{X}_n - \mu_0) \right]$$

and its limiting null distribution can be derived similarly. Below we shall focus our discussion on the case $1/b$ is an integer for simplicity.

In a similar fashion, if we want to construct an MBB-based symmetric confidence interval for μ, we consider the approximation of the sampling distribution of $|\bar{X}_n - \mu|\sqrt{n}$ by the conditional distribution of $|\bar{X}_n^*(l) - \bar{X}_n|\sqrt{n}$ given the data. The corresponding p-value is

$$\begin{aligned}\widetilde{\text{pval}}_{n,l}^{\text{MBB}} &:= E^*[\mathbf{1}\{|\bar{X}_n - \mu_0|\sqrt{n} \leqslant |\bar{X}_n^*(l) - \bar{X}_n|\sqrt{n}\}] \\ &= \frac{1}{N^{R_b}} \sum_{j_1, j_2, \ldots, j_{R_b}=1}^{N} \mathbf{1}\left\{ \left| n^{-1/2} \sum_{h=1}^{R_b} \sum_{s=j_h}^{j_h+l-1} (X_s - \mu_0) - (\bar{X}_n - \mu_0)\sqrt{n} \right| \geqslant |\bar{X}_n - \mu_0|\sqrt{n} \right\}\end{aligned}$$

and it converges in distribution to

$$\tilde{H}(b) = (1-b)^{-R_b} \int_0^{1-b} \ldots \int_0^{1-b} \mathbf{1}\left[\left|\sum_{h=1}^{R_b} \{W(t_h+b) - W(t_h)\} - W(1)\right| \geqslant |W(1)|\right] \mathrm{d}t_1 \ldots \mathrm{d}t_{R_b}$$

under the null. Define

$$\tilde{M}^*_{n,l,B}(x) = \frac{1}{B} \sum_{j=1}^{B} \mathbf{1}\{|\bar{X}^{*,j}_n(l) - \bar{X}_n|\sqrt{n} \leqslant x\}$$

and $\tilde{c}^*_{n,l,B}(1-\alpha) = \inf\{x : \tilde{M}^*_{n,l,B}(x) \geqslant 1-\alpha\}$. Then the fixed-$b$-based $100(1-\alpha)\%$ symmetric confidence interval for μ is $[\bar{X}_n - n^{-1/2}\tilde{c}^*_{n,l,B}\{1-\tilde{H}_\alpha(b)\}, \bar{X}_n + n^{-1/2}\tilde{c}^*_{n,l,B}\{1-\tilde{H}_\alpha(b)\}]$.

3. Finite dimensional parameter

We first introduce some notation. Let $D[0,1]$ be the space of functions on $[0,1]$ which are right continuous and have left limits, endowed with the Skorokhod topology (Billingsley, 1968). Denote by '$\Rightarrow$' weak convergence in $D[0,1]$ or more generally in the $\mathbb{R}^k$-valued function space $D^k[0,1]$, where $k \in \mathbb{N}$. Later in Section 4, we also use '$\Rightarrow$' to denote the weak convergence in $D([-\infty,\infty] \times [0,1])$.

3.1. Subsampling

Following Politis *et al.* (1999a), we assume that the parameter of interest is $\theta(P) \in \mathbb{R}^k$, where P is the joint probability law that governs the P-dimensional stationary sequence $\{X_t\}_{t\in\mathbb{Z}}$. Let $\hat{\theta}_n = \hat{\theta}_n(X_1,\ldots,X_n)$ be an estimator of $\theta = \theta(P)$ based on the observations $(X_1,\ldots,X_n)$. Further we define the subsampling estimator of $\theta(P)$ by $\hat{\theta}_{j,j+l-1} = \hat{\theta}_l(X_j,\ldots,X_{j+l-1})$ on the basis of the subsample $(X_j,\ldots,X_{j+l-1})$, $j = 1,\ldots,N$. Let $\|\cdot\|$ be a norm in $\mathbb{R}^k$. The subsampling-based distribution estimator of $\|\{\hat{\theta}_n - \theta(P)\}\sqrt{n}\|$ is denoted as $\tilde{L}_{n,l}(x) = N^{-1}\sum_{j=1}^{N} \mathbf{1}\{\|(\hat{\theta}_{j,j+l-1} - \hat{\theta}_n)\sqrt{l}\| \leqslant x\}$. In the testing context (say $H_0 : \theta = \theta_0$ *versus* $H_1 : \theta \neq \theta_0$), we define the subsampling-based p-value as

$$\widetilde{\text{pval}}^{\text{SUB}}_{n,l} = N^{-1} \sum_{j=1}^{N} \mathbf{1}\{\|(\hat{\theta}_n - \theta)\sqrt{n}\| \leqslant \|(\hat{\theta}_{j,j+l-1} - \hat{\theta}_n)\sqrt{l}\|\}, \tag{2}$$

where for notational convenience we do not distinguish θ and θ_0 because they are the same under the null hypothesis.

To obtain the limiting null distribution of the p-value under the fixed b framework, we further assume that $\theta(P) = T(F)$, where F is the marginal distribution of $X_1 \in \mathbb{R}^p$, and T is a functional that takes value in $\mathbb{R}^k$. Then a natural estimator of $T(F)$ is $\hat{\theta}_n = T(\rho_{1,n})$, where $\rho_{1,n} = n^{-1}\sum_{t=1}^{n} \delta_{X_t}$ is the empirical distribution. Here δ_x stands for the point mass at x. Similarly, $\hat{\theta}_{j,j+l-1} = T(\rho_{j,j+l-1})$, where $\rho_{j,j+l-1} = l^{-1}\sum_{h=j}^{j+l-1} \delta_{X_h}$. Assume that there is an expansion of $T(\rho_{1,n})$ in the neighbourhood of F, i.e.

$$T(\rho_{1,n}) = T(F) + n^{-1} \sum_{t=1}^{n} \mathrm{IF}(X_t; F) + R_{1,n},$$

where $\mathrm{IF}(X_t; F)$ is the influence function of T (Hampel *et al.*, 1986) defined by

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

and $R_{1,n}$ is the remainder term. Similarly,

$$T(\rho_{j,j+l-1}) = T(F) + l^{-1} \sum_{h=j}^{j+l-1} \mathrm{IF}(X_h; F) + R_{j,j+l-1}.$$

Below are the two key assumptions that we need.

Assumption 1. Assume that $\mathbb{E}\{\mathrm{IF}(X_j; F)\} = 0$ and $n^{-1/2} \Sigma_{j=1}^{\lfloor nr \rfloor} \mathrm{IF}(X_j; F) \Rightarrow \Sigma(P)^{1/2} W_k(r)$, where $\Sigma(P)$ is a positive definite matrix and $W_k(\cdot)$ denotes the kth-dimensional vector of independent Brownian motions.

Assumption 2. Assume that $\|R_{1,n}\| \sqrt{n} = o_p(1)$ and $\sup_{j=1,\ldots,N} \|R_{j,j+l-1}\| \sqrt{l} = o_p(1)$.

Note that assumption 1 is also assumption 1 in Shao (2010a) and its verification has been discussed in remark 1 therein. Assumption 2 is to ensure the negligibility of remainder terms. In the sample mean case, $\mathrm{IF}(X_t; F) = X_t - \mu$ and the remainder terms vanish, so assumption 2 is automatically satisfied and assumption 1 reduces to a functional central limit theorem for the partial sum process of X_t.

Theorem 1. Suppose that assumptions 1 and 2 hold and $b \in (0, 1]$ is held fixed as $n \to \infty$. The limiting null distribution of $\widetilde{\mathrm{pval}}_{n,l}^{\mathrm{SUB}}$ is the distribution of $\tilde{G}(b; k)$, where

$$\tilde{G}(b; k) = (1-b)^{-1} \int_0^{1-b} \mathbf{1}[\|\Sigma(P)^{1/2} W_k(1)\| \leqslant \|\Sigma(P)^{1/2} \{W_k(b+t) - W_k(t) - b\, W_k(1)\}\| / \sqrt{b}]\, \mathrm{d}t.$$

In the special case $k = 1$, $\tilde{G}(b; 1) = \tilde{G}(b)$.

Thus, for a scalar parameter, the limiting null distribution of the p-value is pivotal for a given b and the $100(1-\alpha)\%$ symmetric confidence interval for θ is

$$[\hat{\theta}_n - n^{-1/2} \tilde{c}_{n,l}\{1 - \tilde{G}_\alpha(b)\}, \hat{\theta}_n + n^{-1/2} \tilde{c}_{n,l}\{1 - \tilde{G}_\alpha(b)\}],$$

which reduces to equation (1) in the mean case. To conduct inference for case $k = 1$, we need to know $G_\alpha(b)$, $\tilde{G}_\alpha(b)$, $H_\alpha(b)$ and $\tilde{H}_\alpha(b)$. Following the practice of Kiefer and Vogelsang (2005), we first generate the simulated values for $\alpha = 0.05, 0.1$ and $b = 0.01, 0.02, \ldots, 0.2$; then we fit the quadratic equation $\mathrm{cv}(b) = a_0 + a_1 b + a_2 b^2$ to the simulated values by ordinary least squares. The intercept a_0 was set to be equal to α, so that $\mathrm{cv}(0) = \alpha$. Table 1 reports the estimated

Table 1. Simulated values of $G_\alpha(b)$, $\tilde{G}_\alpha(b)$, $H_\alpha(b)$ and $\tilde{H}_\alpha(b)$ when fitted with a quadratic polynomial $\mathrm{cv}(b) = a_0 + a_1 b + a_2 b^2$, $b \in (0, 0.2]$†

	a_0	a_1	a_2	R^2
$G_{0.05}(b)$	0.05	−0.2289	−0.1325	0.9980
$G_{0.1}(b)$	0.1	−0.1039	−0.8407	0.9997
$\tilde{G}_{0.05}(b)$	0.05	−0.3929	0.6394	0.9978
$\tilde{G}_{0.1}(b)$	0.1	−0.3285	−0.4088	0.9994
$H_{0.05}(b)$	0.05	−0.3431	0.5766	0.9868
$H_{0.1}(b)$	0.1	−0.4079	0.2256	0.9681
$\tilde{H}_{0.05}(b)$	0.05	−0.2121	0.2624	0.9610
$\tilde{H}_{0.1}(b)$	0.1	−0.2461	0.1174	0.9584

†$\alpha = 0.05, 0.1$. The simulated values are based on $n = 5000$ and $n = 50000$ replications. In the MBB case, we use 50000 bootstrap replications.

coefficients and R^2 from the regressions (ranging from 0.9584 to 0.9997), which suggests quite satisfactory fits. For $H_\alpha(b)$ and $\tilde{H}_\alpha(b)$, fitting higher order polynomials does not lead to substantial higher R^2. To simulate $G_\alpha(b)$ and $\tilde{G}_\alpha(b)$ for a given α and $b \in (0, 0.2)$, we generate 5000 IID $N(0,1)$ random variables and use its normalized partial sum to approximate standard Brownian motion. For $H_\alpha(b)$ and $\tilde{H}_\alpha(b)$, we approximate $\mathbb{E}^*$ in the definition of the p-value by performing a bootstrap 50 000 times. 50 000 Monte Carlo replications were used for all the cases. For small α (say $\alpha = 0.01$) and relatively large b, say $b = 0.15, \ldots, 0.2$, our simulated critical values are mostly 0, so we cannot provide a fitted quadratic equation when α is very small. Nevertheless, if the goal is to construct a 90% or 95% symmetric confidence interval, or a 90% equal-tailed confidence interval, or a one-sided confidence interval of nominal coverage 90% or 95%, Table 1 is useful when $b \in (0, 0.2]$.

For a vector parameter (i.e. $k \geqslant 2$), the limiting null distribution of the p-value depends on the unknown long-run variance matrix, so it is not pivotal in general. One way out is to approximate the limiting null distribution by subsampling. Denote by n' the subsampling width at the second stage. Let $l' = \lceil n'b \rceil$ and $N' = n' - l' + 1$. For each subsample $\{X_t, \ldots, X_{t+n'-1}\}$, we define the subsampling counterpart of $\widetilde{\text{pval}}_{n,l}^{\text{SUB}}$ as

$$q_{n',t} = (N')^{-1} \sum_{j=t}^{t+N'-1} \mathbf{1}\{\|(\hat{\theta}_{j,j+l'-1} - \hat{\theta}_{t,t+n'-1})\sqrt{l'}\| \geqslant \|(\hat{\theta}_{t,t+n'-1} - \hat{\theta}_n)\sqrt{n'}\|\}$$

for $t = 1, \ldots, n - n' + 1$. Denote the empirical distribution function of $\{q_{n',t}\}_{t=1}^{n-n'+1}$ by

$$Q_{n,n'}(x) = (n - n' + 1)^{-1} \sum_{j=1}^{n-n'+1} \mathbf{1}(q_{n',j} \leqslant x),$$

which can be used to approximate the sampling distribution or the limiting null distribution of $\widetilde{\text{pval}}_{n,l}^{\text{SUB}}$. Let $c_{n,n',l}(1-\alpha) = \inf\{x : Q_{n,n'}(x) \geqslant 1 - \alpha\}$. Then the calibrated $100(1-\alpha)\%$ subsampling-based confidence region for θ is

$$\{\theta \in \mathbb{R}^k : \widetilde{\text{pval}}_{n,l}^{\text{SUB}} \text{ in equation (2) } \geqslant c_{n,n',l}(\alpha)\}, \tag{3}$$

whereas the traditional subsampling-based confidence region is $\{\theta \in \mathbb{R}^k : \widetilde{\text{pval}}_{n,l}^{\text{SUB}}$ in equation (2) $\geqslant \alpha\}$.

Theorem 2. Assume that $1/n' + n'/n = o(1)$ and $b \in (0, 1]$ is fixed. Suppose that the process X_t is α mixing and $\tilde{G}(b;k)$ is a continuous random variable. Then we have

$$\sup_{x \in \mathbb{R}} |Q_{n,n'}(x) - P\{\tilde{G}(b;k) \leqslant x\}| = o_p(1).$$

Consequently, the asymptotic coverage probability of the confidence region in expression (3) is $1 - \alpha$.

Remark 1. As we have done subsampling twice, this procedure is naturally called double subsampling in the spirit of the double bootstrap. The use of subsampling at the second stage is mainly to approximate the sampling distribution or the limiting null distribution of the p-value, which is unknown under the fixed b asymptotic framework. Of course, the approximation error depends on the subsampling window size n' at the second stage. If we view n'/n as a fixed constant in the above asymptotics, then the asymptotic coverage of the calibrated confidence region is still different from the nominal level. One can perform further calibration by subsampling, which leads to the iterative subsampling, similar to the iterative bootstrap in Beran (1987, 1988). In practice, however, the selection of the subsampling window size at each stage usually involves quite expensive computation, and the (finite sample) improvement in coverage errors is not guaranteed by doing subsampling iteratively.

As pointed out by a referee, a possible alternative approach is to simulate the asymptotic null distribution of the p-value, i.e. the distribution of $\tilde{G}(b;k)$ after plugging in a consistent estimator of the long-run variance matrix. In general, consistent estimation of the long-run variance matrix also involves the bandwidth selection; see for example Politis (2011). Since the above-mentioned double-subsampling approach is also applicable to the infinitely dimensional case (see Section 4), we shall not pursue this alternative approach.

3.2. Moving block bootstrap

For the MBB, we approximate the sampling distribution of $\|(\hat{\theta}_n-\theta)\sqrt{n}\|$ by the conditional distribution of $(\hat{\theta}_n^*-\hat{\theta}_n)\sqrt{n}$, where $\hat{\theta}_n^*=\hat{\theta}_n\{X_1^*(l),\ldots,X_n^*(l)\}$. Define the p-value as $\widetilde{\mathrm{pval}}_{n,l}^{\mathrm{MBB}}:=\mathbb{E}^*\{\mathbf{1}(\|(\hat{\theta}_n-\theta_0)\sqrt{n}\|\leqslant\|(\hat{\theta}_n^*-\hat{\theta}_n)\sqrt{n}\|)\}$. It can be expected that, under certain regularity conditions, the limiting null distribution of $\widetilde{\mathrm{pval}}_{n,l}^{\mathrm{MBB}}$ is the distribution of

$$\tilde{H}(b;k)=\frac{1}{(1-b)^{R_b}}\int_0^{1-b}\ldots\int_0^{1-b}\mathbf{1}\left(\left\|\Sigma(P)^{1/2}\left[\sum_{h=1}^{R_b}\{W_k(t_h+b)-W_k(t_h)\}-W_k(1)\right]\right)\right\| \geqslant\|\Sigma(P)^{1/2}W_k(1)\|\Big]\,\mathrm{d}t_1\ldots\mathrm{d}t_{R_b},$$

which coincides with $\tilde{H}(b)$ when $k=1$. When $k\geqslant 2$, the p-value is not asymptotically pivotal under the fixed b asymptotics, and its sampling distribution can be approximated by subsampling or the MBB. Since the idea is similar to the double-subsampling procedure described above, we omit the details. We mention in passing that Lee and Lai (2009) have recently studied the benefit of performing the double-block bootstrap for the smooth function model.

The p-value-based calibration is closely related to the prepivoting method that was proposed by Beran (1987, 1988). The p-value of a statistic is itself a statistic that has a pivotal limiting distribution or tends to be more pivotal than the original (un-Studentized) statistic. In Beran (1987), the limiting null distribution of the p-value was assumed to be $U(0,1)$, and he focused on the refinement of the approximation error of the sampling distribution of the p-value to $U(0,1)$ by prepivoting and iterative bootstrapping. His treatment is quite general but is mainly focused on the IID data setting. By contrast, we deal with time series with independent data as a special case and the limiting null distribution of the p-value (under the fixed b asymptotics) is not $U(0,1)$. In addition, our calibration can be applied to the inference of infinite dimensional parameters (see Section 4), which was not covered by Beran (1987, 1988). Another related calibration method in the bootstrap literature was proposed by Loh (1987, 1991), who calibrated confidence coefficients by using a consistent estimate of actual coverage probability. For a given confidence interval, its estimated coverage probability is used to alter the nominal level of the interval, and it is shown that the calibrated interval is asymptotically robust under IID assumptions and some regularity conditions. Similar to Beran's work, Loh's discussion is limited to the IID setting and his calibration method seems only applicable to the inference of finite dimensional parameters. For a comprehensive account of bootstrap iteration and calibration, see Hall (1992).

For a finite dimensional parameter, another way of making the statistic more pivotal is to do Studentization using a consistent estimate of the asymptotic variance of the original statistic. For dependent data, this typically involves the estimation of long-run variance by using the lag window type estimate. Although theoretically possible, consistent estimation can be difficult to carry out in practice for some statistics. For example, if $k=p=1$, $\theta=\mathrm{median}(F)$ and $\hat{\theta}_n=\mathrm{median}(X_1,\ldots,X_n)$, then

$$\Sigma(P) = \{4g^2(\theta)\}^{-1} \sum_{k=-\infty}^{\infty} \mathrm{cov}\{1 - 2\mathbf{1}(X_0 \leqslant \theta), 1 - 2\mathbf{1}(X_k \leqslant \theta)\}$$

with $g(\cdot)$ being the density function of X_1. Consistent estimation of $\Sigma(P)$ involves kernel density estimation for $g(\theta)$ and long-run variance estimation for the transformed series $1 - 2\mathbf{1}(X_t \leqslant \theta)$, both of which involve the choice of a bandwidth parameter. By contrast, subsampling and the MBB can be used to provide a consistent variance estimate, which lead to a Studentized statistic, or a p-value, which is more pivotal than the original un-Studentized statistic. Both methods are relatively easier to implement, although they also require the user to choose the subsampling window width or block size. The self-normalized approach of Shao (2010a), which uses recursive subsample estimates in its Studentization, would be another good candidate when a direct consistent long-run variance estimation is difficult, although there is a loss of efficiency under certain loss functions.

4. Infinite dimensional parameter

In previous sections, our discussion focuses on the inference of a finite dimensional parameter, for which a root-n-consistent estimator exists and asymptotic normality holds. In general, the use of subsampling and block bootstrap methods are not limited to the inference for finite dimensional parameters. In the time series setting, they have been used to provide an approximation of the non-pivotal limiting distribution when the parameter of interest is of infinite dimension. Below we shall focus on confidence band construction of the marginal distribution function. A similar development for the inference of normalized spectral distribution functions of a stationary time series is presented in Shao and Politis (2011). In what follows, we use $\|F - G\|_\infty$ to denote $\sup_{x \in \mathbb{R}} |F(x) - G(x)|$.

Consider a stationary sequence $\{X_k, k \in \mathbb{Z}\}$ and let $m(s) = P(X_0 \leqslant s)$ be its marginal distribution. Given observations $\{X_t\}_{t=1}^n$, the empirical process is defined as $m_n(s) = n^{-1} \sum_{k=1}^n \mathbf{1}(X_k \leqslant s)$. More generally, we define the standardized recursive process

$$K_n(s, \lfloor nt \rfloor) = n^{-1/2} \sum_{k=1}^{\lfloor nt \rfloor} \{\mathbf{1}(X_k \leqslant s) - m(s)\}, \qquad t \in [0, 1].$$

Under certain regularity conditions (see Berkes *et al.* (2009)), we have that

$$K_n(s, \lfloor nt \rfloor) \Rightarrow K(s, t). \tag{4}$$

Here $K(s, t)$, $(s, t) \in [-\infty, \infty] \times [0, 1]$ is a two-parameter mean 0 Gaussian process with

$$\mathrm{cov}\{K(s, t), K(s', t')\} = (t \wedge t')\,\Gamma(s, s'),$$

where

$$\Gamma(s, s') = \sum_{k=-\infty}^{\infty} \mathrm{cov}\{\mathbf{1}(X_0 \leqslant s), \mathbf{1}(X_k \leqslant s')\}.$$

To construct a confidence band for $m(\cdot)$, we note that, by the continuous mapping theorem, expression (4) implies that $\|m_n - m\|_\infty \sqrt{n} \to_{\mathrm{D}} \sup_{s \in \mathbb{R}} |K(s, 1)|$. Since $K(s, 1)$ is a Gaussian process with mean 0 and unknown covariance $\mathrm{cov}\{K(s, 1), K(s', 1)\} = \Gamma(s, s')$, direct inference of $m(\cdot)$ is difficult. To circumvent the difficulty, both the MBB and subsampling have been proposed to approximate the limiting distribution $\sup_{s \in \mathbb{R}} |K(s, 1)|$ consistently; see Bühlmann (1994), Naik-Nimbalkar and Rajarshi (1994) and Politis *et al.* (1999b). Below we shall focus our discussion on the subsampling method, and a similar argument applies to the MBB approach in view of the argument in Section 2.2. Let $g_n(t, s) = \{m_{t, t+l-1}(s) - m_n(s)\}\sqrt{l}$, $t = 1, \ldots, N = n - l + 1$, be the subsampling counterpart of $\{m_n(s)-m(s)\}\sqrt{n}$, where $m_{t,t+l-1}(s)=l^{-1}\Sigma_{h=t}^{t+l-1}\mathbf{1}(X_h\leqslant s)$. Assuming that $l/n+1/l=o(1)$ and other regularity conditions, Politis *et al.* (1999b) showed that the subsampling approximation based on $\{g_n(t,s)\}_{t=1}^N$ is consistent in certain function spaces. This implies that the sampling distribution of $\|m_n-m\|_\infty\sqrt{n}$ (or the distribution of $\sup_{s\in\mathbb{R}}|K(s,1)|$) can be consistently approximated by the empirical distribution of $\{\|m_{t,t+l-1}-m_n\|_\infty\}_{t=1}^N\sqrt{l}$. The above result is obtained under the small b asymptotics. To introduce our calibration method, we again start with the p-value and study its limiting null distribution under the fixed b asymptotics. For notational simplicity, we do not distinguish the true marginal distribution function $m(x)$ and the hypothesized function $m_0(x)$, because they are identical under the null hypothesis.

Define the p-value

$$\mathrm{pval}_{n,l}^{\mathrm{E}}=N^{-1}\sum_{j=1}^{N}\mathbf{1}(\|m_{j,j+l-1}-m_n\|_\infty\sqrt{l}\geqslant\|m_n-m\|_\infty\sqrt{n}).\tag{5}$$

Let $b=l/n$. Under the fixed b asymptotics, the limiting null distribution of the p-value is the distribution of $J(b)$, where

$$J(b):=(1-b)^{-1}\int_0^{1-b}\mathbf{1}\left\{\sup_{s\in\mathbb{R}}|K(s,r+b)-K(s,r)-b\,K(s,1)|/\sqrt{b}\geqslant\sup_{s\in\mathbb{R}}|K(s,1)|\right\}\mathrm{d}r.$$

The distribution of $J(b)$ is not pivotal for a given b, because it depends on the Gaussian process $K(s,t)$, whose covariance structure is tied to the unknown dependence structure of X_t. So subsampling at the first stage is insufficient under the fixed b asymptotic framework. It is worth noting that, in the IID data setting, the quantity $\|m_n-m\|_\infty\sqrt{n}$ is pivotal provided that m is continuous, so the inferential difficulty is mainly caused by the presence of unknown weak dependence.

To make the inference feasible, we propose to approximate the sampling distribution of the p-value or its limiting null distribution by subsampling; see Section 3.1. Let n' be the subsampling window size at the second stage, $l'=\lceil n'b\rceil$ and $N'=n'-l'+1$. For each subsample $\{X_t,\ldots,X_{t+n'-1}\}$, the subsampling counterpart of $\mathrm{pval}_{n,l}^{\mathrm{E}}$ is defined as

$$h_{n',t}=(N')^{-1}\sum_{j=t}^{t+N'-1}\mathbf{1}(\|m_{j,j+l'-1}-m_{t,t+n'-1}\|_\infty\sqrt{l'}\geqslant\|m_{t,t+n'-1}-m_n\|_\infty\sqrt{n'})$$

for $t=1,\ldots,n-n'+1$. Then we can approximate the sampling distribution of $\mathrm{pval}_{n,l}^{\mathrm{E}}$ or its limit null distribution $J(b)$ by the empirical distribution that is associated with $\{h_{n',t}\}_{t=1}^{n-n'+1}$, which is denoted

$$J_{n,n'}(x)=(n-n'+1)^{-1}\sum_{t=1}^{n-n'+1}\mathbf{1}(h_{n',t}\leqslant x).$$

For a given $\alpha\in(0,1)$, the $100(1-\alpha)\%$ traditional subsampling-based confidence band for $m(\cdot)$ is

$$\{m:m\text{ is a distribution function and }\mathrm{pval}_{n,l}^{\mathrm{E}}\text{ in equation (5)}\geqslant\alpha\},$$

and the calibrated confidence band is

$$\{m:m\text{ is a distribution function and }\mathrm{pval}_{n,l}^{\mathrm{E}}\text{ in equation (5)}\geqslant\bar{c}_{n,n',l}(\alpha)\},\tag{6}$$

where $\bar{c}_{n,n',l}(1-\alpha)=\inf\{x:J_{n,n'}(x)\geqslant 1-\alpha\}$. The following theorem states the consistency of subsampling at the second stage, which implies that the coverage for the calibrated confidence band is asymptotically correct. Let

$$\tilde{V}_b(r,\varepsilon) := P\left\{\left|\sup_{s\in\mathbb{R}} |K(s,r+b) - K(s,r) - b\,K(s,1)|/\sqrt{b} - \sup_{s\in\mathbb{R}} |K(s,1)|\right| = \varepsilon\right\}.$$

Theorem 3. Assume that $1/n' + n'/n = o(1)$, equation (4) and $b \in (0,1]$ is fixed.

(a) The limiting null distribution of the p-value in equation (5) is the distribution of $J(b)$ provided that $\tilde{V}_b(r,0) = 0$ for every $r \in [0, 1-b]$.
(b) Suppose that the process X_t is α mixing, $J(b)$ is a continuous random variable and $\tilde{V}_b(r,\varepsilon) = 0$ for every $r \in [0, 1-b]$ and $\varepsilon \geqslant 0$. Then we have

$$\sup_{x\in\mathbb{R}} |J_{n,n'}(x) - P\{J(b) \leqslant x\}| = o_p(1).$$

Consequently, the asymptotic coverage probability of the confidence band in expression (6) is $1-\alpha$.

The conditions on $J(b)$ and $\tilde{V}_b(r,\varepsilon)$ are technical ones that are not easy to verify. The verification seems related to the regularity of the distribution of the maximum of Gaussian processes; see Diebolt and Posse (1996), Azaïs and Wschebor (2001) and references therein. We conjecture that they hold for a large class of Gaussian processes. Note that our calibration is based on the subsampling-based approximation to the sampling distribution of the p-value, which is obtained by doing the subsampling in the first stage. As the p-value is a prepivoted statistic, we are effectively combining the prepivoting idea with subsampling in the infinite dimensional parameter case, for which the usual Studentizing technique in the finite dimensional parameter case does not seem to apply. The idea of prepivoting (using the p-value) in the infinite dimensional parameter case seems new and quite general. We can also use the MBB in the first stage to obtain a p-value or in the second stage to approximate the sampling distribution of the p-value. But the implementation of the MBB in this setting seems very computationally demanding, especially when the block size is chosen through some data-driven algorithms. For this reason, we shall focus on the subsampling method in simulation studies for the infinite dimensional case.

5. Simulation results

In this section, we conduct simulation studies to evaluate the accuracy of the asymptotic approximations that are provided by both small b and fixed b approaches to the finite sample distribution. Specifically, we examine the empirical coverage probabilities and the volumes of confidence sets to see whether the fixed b approach corresponds to smaller coverage errors.

5.1. *Finite sample performance of confidence intervals*

In this subsection, we consider a univariate stationary time series model with various types of dependence structure. To be specific, we let

$$X_t = \mu + u_t,$$

$$u_t = \rho u_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}, \qquad \varepsilon_t \sim \text{IID } N(0,1).$$

We consider

(a) AR(1)–$N(0,1)$ error, $(\rho,\theta) = (0,0)$, $(0.5,0)$, $(0.8,0)$,
(b) moving average MA(1)–$N(0,1)$ error, $(\rho,\theta) = (0, -0.5)$,

and their corresponding AR(1)–EXP(1) and MA(1)–EXP(1) models, where $\varepsilon_t \sim$ IID EXP$(1) - 1$ has mean 0 and unit variance but with an asymmetric distribution. Following the suggestion of a referee, we also include two non-linear time series models: non-linear 1, $X_t = 0.6 \sin(X_{t-1}) + \varepsilon_t$, where $\varepsilon_t \sim \text{IID}\, N(0,1)$ (this model was used in the simulation work of Paparoditis and Politis (2001) and Shao (2010c)); non-linear 2 (a threshold auto-regressive model of order 1),

$$X_t = 0.3X_{t-1}\,\mathbf{1}(X_{t-1} > 0) + 0.8X_{t-1}\,\mathbf{1}(X_{t-1} \leqslant 0) + \varepsilon_t,$$

where $\varepsilon_t \sim \text{IID}\, N(0,1)$. The sample size $n = 100$ and the number of bootstrap replications is 5000. The bandwidth parameter l varies from 3 to 16, i.e. $b = 0.03, 0.04, \ldots, 0.16$. We examine the empirical coverages and average widths of symmetric confidence intervals for $\mu = \mathbb{E}(X_1)$ and the 25% trimmed mean based on 10000 replications. The nominal coverage is set to be 95%.

For the models with normally distributed errors, the results for the mean case are depicted in Fig. 1, in which Figs 1(a)–1(d) show the empirical coverages and Figs 1(e)–1(h) show the corresponding ratios of average interval widths (fixed b over small b). For both subsampling and the MBB, undercoverage occurs and it becomes more severe as the dependence strengthens. The empirical coverages for the fixed b approach are closer to the nominal level than those for the small b approach, with the difference between two empirical coverages increasing as b becomes large. In contrast, the fixed-b-based interval is slightly wider than its small b counterpart, with the ratio of widths increasing with respect to b in general. These findings are consistent with the intuition that, the larger b is, the more accurate the fixed-b-based approximation provides relative to its small b counterpart. The intervals that are constructed on the basis of the MBB have noticeably better coverage than those based on subsampling, especially for large b. For the MA(1) model with $\theta = -0.5$, there is overcoverage for the fixed-b-based interval, which is usually slightly more conservative than the small b counterpart. The overcoverage in the case of negatively correlated time series corresponds to the underrejection for Kiefer and Vogelsang's (2005) Studentized statistic when using the normal approximation (i.e. the small b approach) and b is small (see Fig. 1 therein), so our results are in a sense consistent with those in Kiefer and Vogelsang (2005). Practically speaking, overcoverage is less harmful to the practitioner than undercoverage, so we are less concerned in practice.

The results for the models with exponentially distributed errors, the non-linear models and the trimmed mean case are not included here but are presented in Shao and Politis (2011). Fig. 2 in Shao and Politis (2011) shows that the results corresponding to exponentially distributed errors are very similar to those for normally distributed errors, suggesting that the asymmetric shape of exponentially distributed errors has little effect on the finite sample performance. Figs 3 and 4 in Shao and Politis (2011) present the results for the trimmed mean case, which are fairly similar to the results in the mean case. Additionally, the results for the non-linear models as presented in Fig. 5 of Shao and Politis (2011) resemble those for AR(1)–N(0,1) models with $\rho = 0.5$ in both the mean and the trimmed mean case, indicating that non-linearity does not affect our results much.

Owing to the duality of confidence interval construction and hypothesis testing, we would expect that the fixed b approach leads to better size (i.e. size closer to the nominal size) in all the models except MA(1) with $\theta = -0.5$, at the sacrifice of (raw) power. The power loss is expected to be moderate because the ratio of fixed-b-based interval width over the small-b-based interval width is quite close to 1. Overall, the simulation results demonstrate that the fixed b approach delivers more accurate inference for both subsampling and the MBB in most situations owing to its more accurate approximation to the finite sample distribution. Of course, we show only the improved accuracy of the fixed b approximation for a specific $\alpha = 0.05$, which is also what Kiefer and Vogelsang (2005) did. We also tried $\alpha = 10\%$ and qualitatively similar results are obtained. It would be interesting to provide some theoretical justifications on the order of

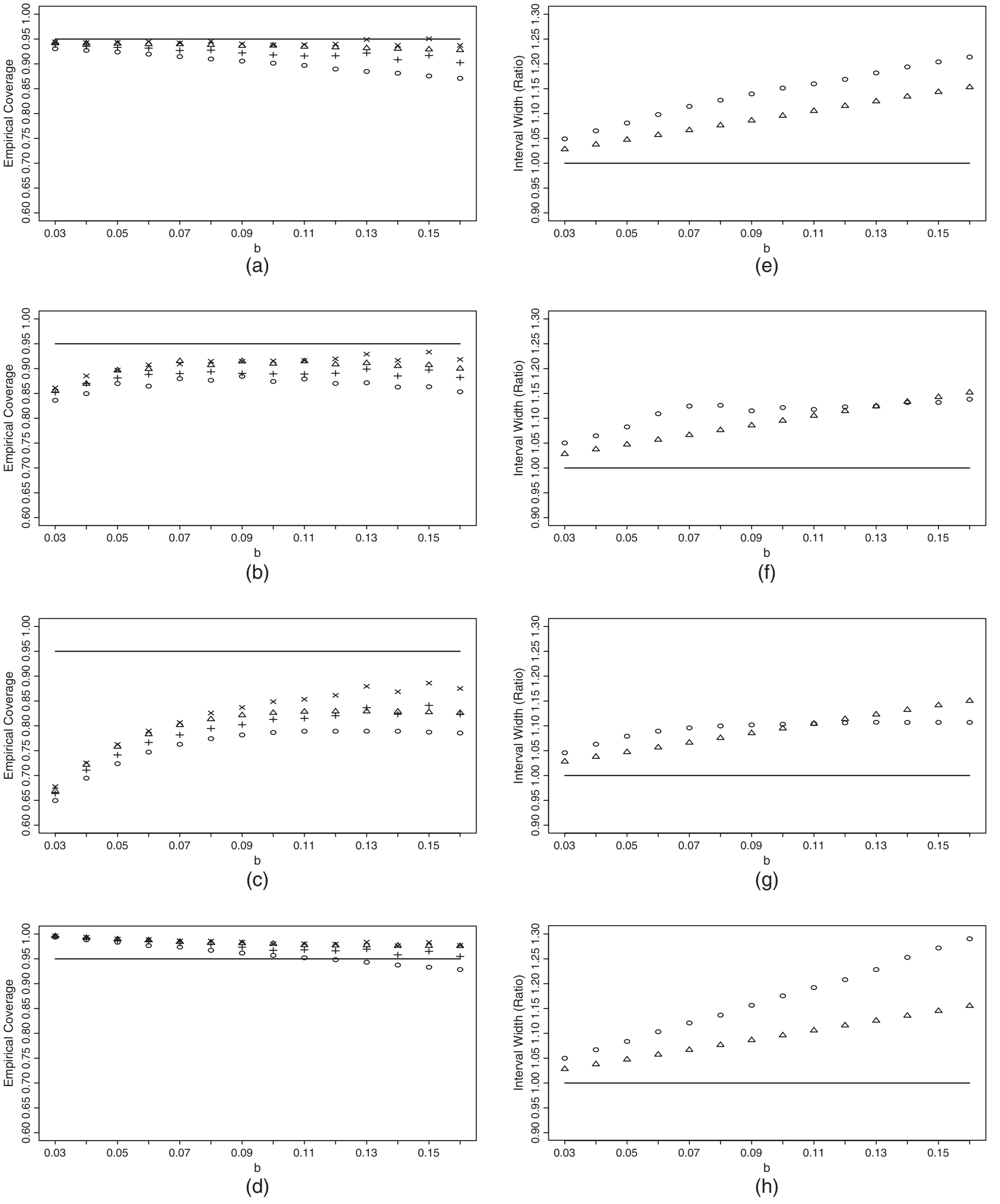

Fig. 1. (a)–(d) Empirical coverage probabilities (○, traditional, subsampling; Δ, calibrated, subsampling; +, traditional, MBB; ×, calibrated, MBB) and (e)–(h) ratios of interval widths (calibrated fixed *b* over traditional small *b*; ○, subsampling; Δ, MBB) for the mean and for the models with normally distributed errors (sample size $n = 100$; number of replications 10000): (a), (e) $\rho = 0$; (b), (f) $\rho = 0.5$; (c), (g) $\rho = 0.8$; (d), (h) $\theta = -0.5$

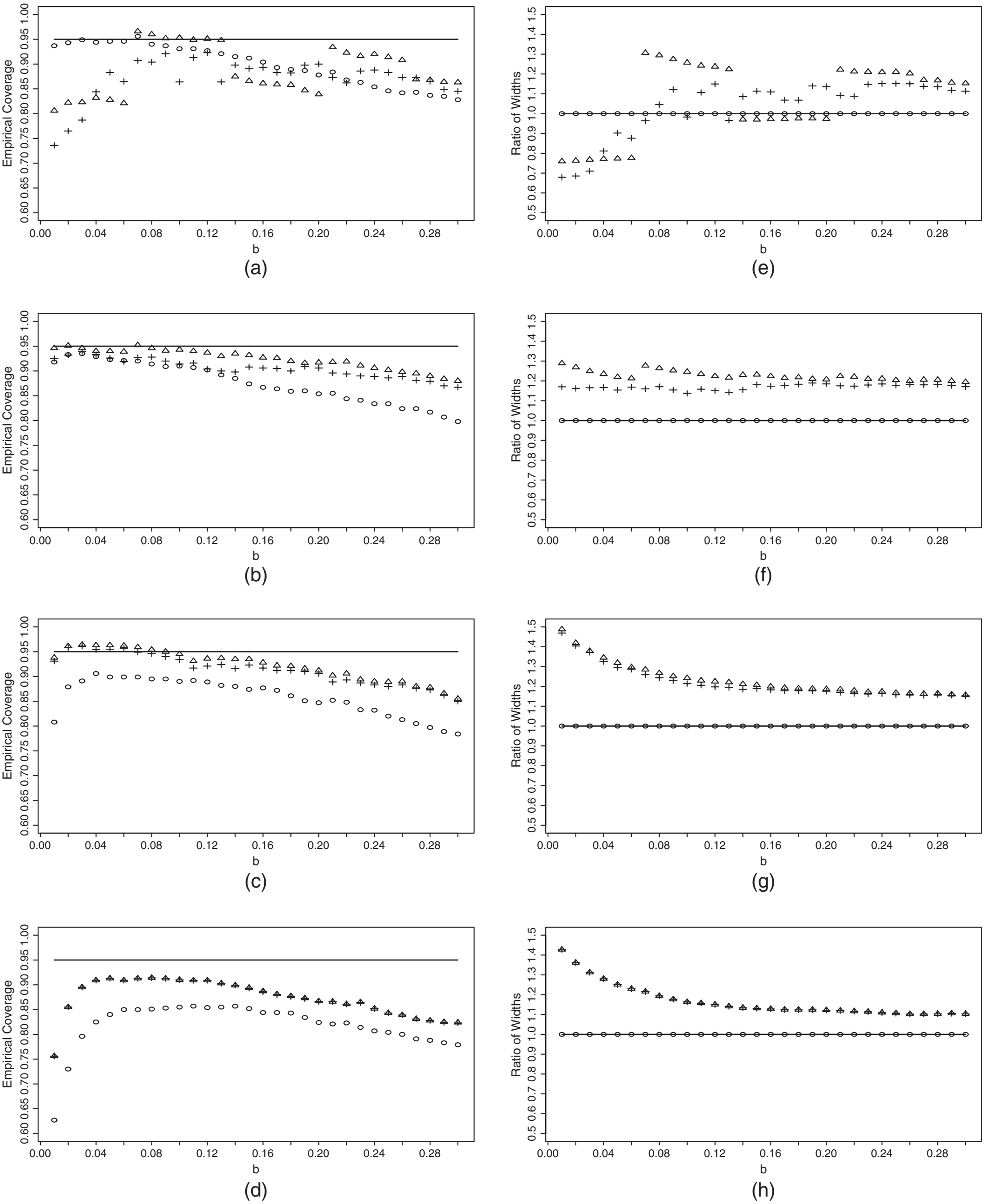

Fig. 2. (a)–(d) Empirical coverage probabilities and (e)–(h) ratios of radii of confidence regions (calibrated fixed b over traditional small b) for the vector parameter and for the models with normally distributed errors (sample size $n = 200$; number of replications 1000; ○, traditional; Δ, calibrated (fixed); +, calibrated (data driven)): (a), (e) $\rho = -0.6$; (b), (f) $\rho = 0$; (c), (g) $\rho = 0.5$; (d), (h) $\rho = 0.8$

the error rejection probability. For the subsampling method, this boils down to the order of $\sup_{\alpha\in[0,1]}|P(\text{pval}_{n,l}^{\text{SUB}}\leqslant\alpha)-P\{\tilde{G}(l/n)\leqslant\alpha\}|$ under the fixed b framework. Under the small b framework, the error is $\sup_{\alpha\in[0,1]}|P(\text{pval}_{n,l}^{\text{SUB}}\leqslant\alpha)-\alpha|$, which is expected to be larger. A formal theoretical proof seems quite challenging and is left for future research.

5.2. Finite sample performance of confidence regions and confidence bands

In this subsection, we examine the coverage probabilities of confidence regions for the vector parameter of the mean and median, and confidence bands for the marginal distribution function $m(\cdot)$. Let $\{X_t\}_{t=1}^n$ be generated from the AR(1) model: $X_t=\rho X_{t-1}+e_t$, where $\rho=-0.6, 0, 0.5, 0.8$, $e_t\sim\text{IID}\,N(0,1)$ or $\text{EXP}(1)-1$. The sample size $n=200$ and the number of replications is 1000. We use the Euclidean norm in the confidence region construction. For both confidence regions and confidence bands, we compared the following three schemes:

(a) the traditional subsampling-based confidence region (band);
(b) the calibrated subsampling-based confidence region (band) with a fixed n', where $n'=15$ for confidence region construction and $n'=30$ for confidence band construction;
(c) the calibrated subsampling-based confidence region (band) with n' chosen in a data-driven fashion.

Here we employ a variant of a block size selection procedure that was proposed in Bickel and Sakov (2008) for the m-out-of n bootstrap (also see Götze and Račkauskas (2001)), which is closely related to the subsampling method. The use of Bickel and Sakov's automatic bandwidth selection in the subsampling context has been explored in Jach *et al.* (2011) recently. The procedure consists of the following steps (in the confidence band case).

Step 1: for a predetermined interval $[K_1,K_2]$ and $g\in(0,1)$, we consider a sequence of n_js of the form $n_j=\lfloor g^{j-1}K_2\rfloor$, for $j=1,2,\ldots,\lfloor\log(K_2/K_1)/\{-\log(g)\}\rfloor$.
Step 2: for each n_j, find J_{n,n_j}, where J_{n,n_j} is the subsampling-based distribution estimator for the sampling distribution of the p-value.
Step 3: set $j_0=\arg\min_{j=1,\ldots,\lfloor\log(K_2/K_1)/\{-\log(g)\}\rfloor}\sup_{x\in\mathbb{R}}|J_{n,n_j}(x)-J_{n,n_{j+1}}(x)|$. Then the optimal block size is $g^{j_0}K_2$. If the difference is minimized for a few values of j, then pick the largest among them.

In our simulation experiment, we set $(K_1,K_2,g)=(5,40,0.75)$ for confidence region construction and $(K_1,K_2,g)=(10,60,0.75)$ for confidence band construction, which correspond to a sequence of block lengths of $(40,30,22,16,12,9,7,5)$ and $(60,45,33,25,18,14,10)$ respectively.

Fig. 2 depicts the empirical coverages and the ratios of the radii of the confidence regions over that delivered by the uncalibrated traditional subsampling-based region for the vector parameter and for the models with normally distributed errors. The symbols 'traditional', 'calibrated (fixed)' and 'calibrated (data driven)' correspond to schemes (a)–(c) that were described above. When $\rho=0, 0.5, 0.8$, there is undercoverage associated with the traditional subsampling-based approach and the coverage errors increase with respect to the magnitude of dependence. The improvement in coverage that is offered by the calibration is apparent in these cases and it holds uniformly over the range of bs under examination. In contrast, the corresponding radius of the calibrated region is slightly larger than that of the uncalibrated counterpart. In the case $\rho=-0.6$, the calibrated region performs worse compared with the traditional counterpart when b is small but still offers some improvement when b is large. It is not fully clear why this occurs. Nevertheless, it suggests that caution must be exercised in the use of fixed-b-based calibration

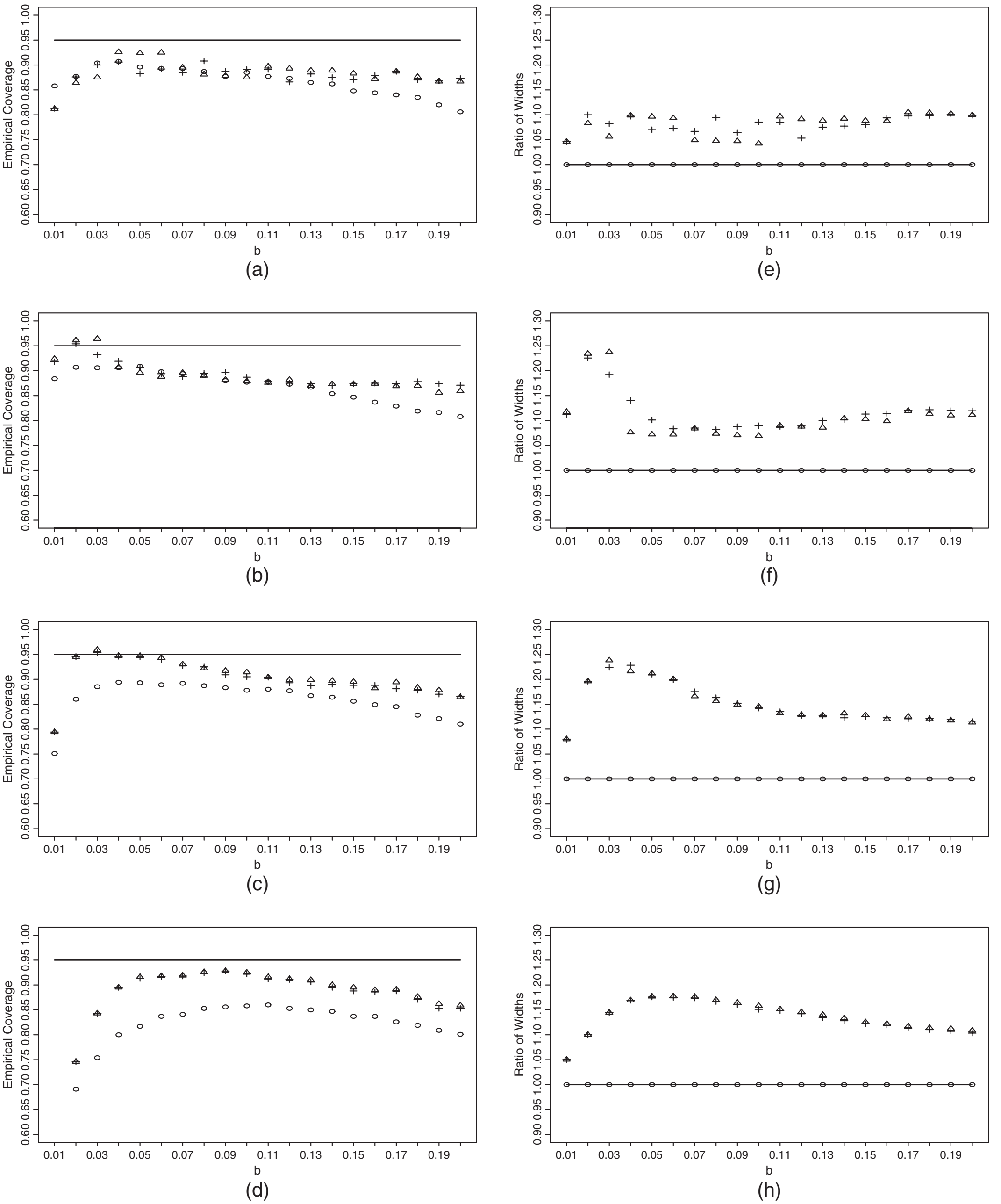

Fig. 3. (a)–(d) Empirical coverage probabilities and (e)–(h) ratios of bandwidths (calibrated fixed b over traditional small b) for the marginal distribution function (sample size $n=200$; number of replications 1000; ○, traditional; Δ, calibrated (fixed); +, calibrated (data driven)): (a), (e) $\rho=-0.6$; (b), (f) $\rho=0$; (c), (g) $\rho=0.5$; (d), (h) $\rho=0.8$

when the auto-correlations of the series have alternating signs. The results for the case of exponentially distributed errors are presented in Fig. 7 of Shao and Politis (2011) and are very close to those for the normally distributed errors.

Fig. 3 has the same format as Fig. 2 and Figs 3(e)–3(h) show the ratio of the mean bandwidth over that delivered by the uncalibrated traditional subsampling-based band. For the marginal distribution function, there is an apparent undercoverage for the traditional subsampling-based confidence band in all cases with coverage errors increasing with respect to the magnitude of dependence (compare the plots for $\rho = 0, 0.5, 0.8$), especially at small bs. When $\rho = 0, 0.5, 0.8$ and for almost all $b = 0.01, \ldots, 0.2$, the coverages that are delivered by the calibrated bands based on fixed or data-driven subsampling widths are closer to the nominal level than the traditional counterpart. When $\rho = -0.6$, the calibrated bands based on the fixed or data-dependent bandwidths improve the coverage when $b \geqslant 0.04$, but fail to do so when $b = 0.01, 0.02, 0.03$, suggesting that potential improvements can be made about the selection of n'. In all cases, the calibrated bands are slightly wider than the uncalibrated counterpart, but the ratios are quite close to 1. The 'better coverage but wider band' phenomenon is in accordance with the 'better coverage but wider interval' finding in the scalar parameter case. The two calibrated bands perform similarly in most situations, and their performance is strikingly close when $\rho = 0.8$. Fig. 9 in Shao and Politis (2011) plots the empirical coverages and ratios of mean bandwidths with respect to $b = 0.04, \ldots, 0.3$ for the normalized spectral distribution; the improvement of the calibration in terms of coverage error is quite substantial when $\rho = -0.6, 0.5, 0.8$. When $\rho = 0$, the calibrated bands are conservative when b is relatively small, but again provide some improvement when b is close to 0.3. Overall the results for the normalized spectral distribution function are qualitatively similar to those for the marginal distribution function. On the basis of simulation results for confidence intervals that were reported in Section 5.1 and for confidence regions and bands reported in this subsection, it appears that the calibration works very effectively when the series is positively dependent.

6. Conclusion

Subsampling and block-based bootstrap methods have been shown to be widely applicable to many inference problems in time series analysis. The fixed b asymptotics that are developed here explicitly capture the choice of bandwidth parameter in subsampling and the MBB and the resulting first-order approximation is expected to be more accurate than that provided by the small b asymptotics. As demonstrated in Section 5, the fixed-b-based calibrated confidence intervals (regions or bands) provide an unambiguous improvement over the uncalibrated counterparts in terms of coverage errors in most cases considered. Our calibration method is developed by estimating the sampling distribution of the p-value, which relates to the prepivoting method of Beran (1987, 1988) and the confidence coefficient calibration method of Loh (1987). However, our proposal differs from theirs in two important respects.

(a) The limiting null distribution of the p-value is not (necessarily) $U(0, 1)$, which is the case for Beran (1987, 1988). In our setting, a pivotal limiting distribution exists in the scalar parameter case, but not in the case of a vector parameter and infinite dimensional parameter, for which the subsampling method is used to provide a good approximation.
(b) Their discussions are limited to the IID data setting and inference for finite dimensional parameters. In contrast, our treatment goes substantially beyond their developments by allowing for time series data and the inference of infinite dimensional parameters.

Coupled with the recently developed fixed b approach (Kiefer and Vogelsang, 2005) in the econometrics literature, we provide a general recipe for the calibration of the traditional resampling-based inference procedures when smoothing parameters, such as window width in subsampling and block size in the MBB, are used to accommodate the dependence.

To conclude the paper, we provide a discussion of open problems and possible extensions.

(a) Our method can be used as a calibration tool for a properly chosen smoothing parameter and it is important practically to choose the smoothing parameter in a sensible way. The choice of subsampling width and block size for the block-based bootstrap has been discussed in chapter 9 of Politis *et al.* (1999a) and chapter 7 of Lahiri (2003). It seems natural to ask whether it is meaningful to consider the optimal smoothing parameter selection from a fixed-b-based viewpoint, as opposed to the small-b-based approach (see for example Bühlmann and Künsch (1999) and Politis and White (2004)). A high order expansion of the sampling distribution of the p-value under the null and alternative hypotheses seems needed to tackle this issue.

(b) The development in this paper is confined to time series, although subsampling and block-based bootstrap methods have been extended to spatial settings (see chapter 5 of Politis *et al.* (1999a) and chapter 12 of Lahiri (2003) and references therein). An extension of the fixed-b-based calibration idea to spatial settings is expected to be possible but seems non-trivial for irregularly spaced spatial data.

(c) In addition, we impose the weak dependence throughout so that the asymptotic normality or functional central limit theorem with root-n-convergence rate hold. When the time series is long range dependent, the subsampling method has been proved to be consistent in some situations (see Hall *et al.* (1998) and Nordman and Lahiri (2005)). It would be interesting to extend the fixed b approach to calibrate the subsampling-based inference in these settings.

(d) A close relative of block-based bootstraps is the so-called sieve bootstrap (Bühlmann, 1997), which also involves a bandwidth parameter (i.e. the order of the approximating auto-regressive model). It is natural to ask whether it is possible to extend the fixed b approach to calibrate the sieve-bootstrap-based confidence sets. We leave these possible extensions for future work.

Acknowledgements

Shao's research is supported in part by National Science Foundation grant DMS-11-04545 and Politis's research is partially supported by National Science Foundation grant DMS-10-07513. The authors thank Xianyang Zhang for helpful comments and Jason Schweinsberg for suggesting some useful references on Gaussian processes. We are also grateful to two referees for constructive comments, which led to substantial improvements.

Appendix A

A.1. Proof of theorem 1

For convenience of notation, let $Y_h = \mathrm{IF}(X_h; F)$ and $\Delta = \Sigma(P)^{1/2}$. Further let $T_{n,j} = \mathbf{1}\{\|(\hat\theta_n - \theta_0)\sqrt{n}\| \leqslant \|(\hat\theta_{j,j+l-1} - \hat\theta_n)\sqrt{l}\|\}$ and

$$\tilde T_{n,j} = \mathbf{1}\left[\left\|n^{-1/2}\sum_{j=1}^{n} Y_j\right\| \leqslant \left\|l^{-1/2}\left\{\sum_{h=j}^{j+l-1} Y_h - (l/n)\sum_{j=1}^{n} Y_j\right\}\right\|\right].$$

Then $\widetilde{\mathrm{pval}}_{n,l}^{\mathrm{SUB}} = N^{-1}\Sigma_{j=1}^{N} T_{n,j}$. Let $D_n(\varepsilon) = \{\|R_{1,n}\sqrt{n}\| < \varepsilon,\ \sup_{j=1,\ldots,N}\|R_{j,j+l-1}\sqrt{l}\| < \varepsilon\}$ for any $\varepsilon > 0$. Then $P\{D_n(\varepsilon)\} \to 1$ as $n \to \infty$. On $D_n(\varepsilon)$, we have that

$$|T_{n,j}-\tilde{T}_{n,j}|\leqslant \mathbf{1}\left[\left|\left\|n^{-1/2}\sum_{j=1}^{n}Y_j\right\|-\left\|l^{-1/2}\left\{\sum_{h=j}^{j+l-1}Y_h-(l/n)\sum_{j=1}^{n}Y_j\right\}\right\|\right|\leqslant 2\varepsilon\right].$$

So the expression $N^{-1}\Sigma_{j=1}^{N}|T_{n,j}-\tilde{T}_{n,j}|$ is bounded by

$$N^{-1}\sum_{j=1}^{N}\mathbf{1}\left[\left|\left\|n^{-1/2}\sum_{j=1}^{n}Y_j\right\|-\left\|l^{-1/2}\left\{\sum_{h=j}^{j+l-1}Y_h-(l/n)\sum_{j=1}^{n}Y_j\right\}\right\|\right|\leqslant 2\varepsilon\right],$$

which, by the continuous mapping theorem, converges in distribution to $I(b,\varepsilon)$, where

$$I(b,\varepsilon):=(1-b)^{-1}\int_0^{1-b}\mathbf{1}[|\|\Delta W_k(1)\|-\|\Delta\{W_k(b+t)-W_k(t)-b\,W_k(1)\}/\sqrt{b}\||\leqslant 2\varepsilon]\,\mathrm{d}t.$$

It is not difficult to see that, for each $t\in[0,1-b]$, the integrand in $I(b,\varepsilon)\downarrow 0$ almost surely as $\varepsilon\downarrow 0$, which implies that $\lim_{\varepsilon\downarrow 0} I(b,\varepsilon)=0$ almost surely by the Lebesgue dominated convergence theorem. Since $N^{-1}\Sigma_{j=1}^{N}\tilde{T}_{n,j}\rightarrow_{\mathrm{D}}\tilde{G}(b;k)$ by the continuous mapping theorem, the conclusion follows by letting $\varepsilon\downarrow 0$ and $n\rightarrow\infty$.

We provide a justification for the use of the continuous mapping theorem above. For any $x\in D^k[0,1]$, define the functional

$$f_1(x)=(1-b)^{-1}\int_0^{1-b}\mathbf{1}[|\|x(1)\|-\|\{x(b+t)-x(t)-b\,x(1)\}/\sqrt{b}\||\leqslant 2\varepsilon]\,\mathrm{d}t$$

and

$$f_2(x)=(1-b)^{-1}\int_0^{1-b}\mathbf{1}[\|x(1)\|\leqslant\|\{x(b+t)-x(t)-b\,x(1)\}/\sqrt{b}\|]\,\mathrm{d}t.$$

To use the continuous mapping theorem, we need to show that both f_1 and f_2 are $\Delta W_k(\cdot)$ continuous almost surely. We shall focus on f_1 and the same argument applies to f_2. Define $D_{f_1}=\{x: f_1$ is not continuous at $x\}$. Then

$$D_{f_1}\subset\tilde{D}_{f_1}=\{x:\lambda\{t\in[0,1-b]:\|x(1)\|-\|\{x(b+t)-x(t)-b\,x(1)\}/\sqrt{b}\|=\pm 2\varepsilon\}>0\},$$

where λ stands for Lebesgue measure. It is enough to show that $P\{\Delta W_k(\cdot)\in\tilde{D}_{f_1}\}=0$. For this, we note that

$$\begin{aligned}\mathbb{E}\left(\int_0^{1-b}\mathbf{1}[\|\Delta W_k(1)\|-\|\{\Delta W_k(b+t)-\Delta W_k(t)-b\,\Delta W_k(1)\}/\sqrt{b}\|=\pm 2\varepsilon]\right)\mathrm{d}t&\\ =\int_0^{1-b}P[\|\Delta W_k(1)\|-\|\{\Delta W_k(b+t)-\Delta W_k(t)-b\,\Delta W_k(1)\}/\sqrt{b}\|=\pm 2\varepsilon]\,\mathrm{d}t&\\ =0&\end{aligned}$$

where we have used the fact that, for each $t\in[0,1-b]$,

$$P[\|\Delta W_k(1)\|-\|\{\Delta W_k(b+t)-\Delta W_k(t)-b\,\Delta W_k(1)\}/\sqrt{b}\|=\pm 2\varepsilon]=0. \quad (7)$$

Fact (7) can be easily shown by noting that the joint distribution of $(\Delta W_k(1),\{\Delta W_k(b+t)-\Delta W_k(t)-b\,\Delta W_k(1)\}/\sqrt{b})$ is multivariate normal with a positive definite covariance matrix. So $P\{\Delta W_k(\cdot)\in\tilde{D}_{f_1}\}=0$ holds and the use of the continuous mapping theorem is justified. The proof is thus complete.

A.2. Proof of theorem 2

The proof of theorem 2 is similar to that of theorem 3, so we omit the details.

A.3. Proof of theorem 3

(a) The proof follows from the use of the continuous mapping theorem. Here the mapping $f: D([-\infty,\infty]\times[0,1])\rightarrow\mathbb{R}$ is defined as

$$f(x) = (1-b)^{-1}\int_0^{1-b} \mathbf{1}\left\{\sup_{s\in\mathbb{R}} |x(s,r+b) - x(s,r) - b\,x(s,1)|/\sqrt{b} \geqslant \sup_{s\in\mathbb{R}} |x(s,1)|\right\} \mathrm{d}r.$$

Following the argument in the proof of theorem 1, we can show that, if $\tilde{V}_b(r,0)=0$ for every $r\in[0,1-b]$, then the mapping f is K continuous, i.e. the probability that the Gaussian process $K(\cdot,\cdot)$ falls into the discontinuity set of f is 0. This completes the proof.

(b) In view of the continuity assumption of $J(b)$ and the monotonicity of $J_{n,n'}(x)$, it suffices to show that $J_{n,n'}(x) = P\{J(b)\leqslant x\} + o_p(1)$ for each $x\in\mathbb{R}$. Let

$$\hat{h}_{n',t} = (N')^{-1} \sum_{j=t}^{t+N'-1} \mathbf{1}(\|m_{j,j+l'-1} - m_{t,t+n'-1}\|_\infty \sqrt{l'} \geqslant \|m_{t,t+n'-1} - m\|_\infty \sqrt{n'})$$

for $t=1,\ldots,n-n'+1$ and

$$\hat{J}_{n,n'}(x) = (n-n'+1)^{-1} \sum_{t=1}^{n-n'+1} \mathbf{1}(\hat{h}_{n',t} \leqslant x).$$

Note that

$$\|m_{t,t+n'-1} - m\|_\infty \sqrt{n'} - \|m_n - m\|_\infty \sqrt{n'} \leqslant \|m_{t,t+n'-1} - m_n\|_\infty \sqrt{n'} \leqslant \|m_{t,t+n'-1} - m\|_\infty \sqrt{n'} + \|m_n - m\|_\infty \sqrt{n'}.$$

For any $\varepsilon>0$, let $E_n(\varepsilon) = \{\|m_n - m\|_\infty \sqrt{n'} \leqslant \varepsilon\}$ and

$$V_b(r,\varepsilon) := P\{|\sup_{s\in\mathbb{R}} |K(s,r+b) - K(s,r) - b\,K(s,1)|/\sqrt{b} - \sup_{s\in\mathbb{R}} |K(s,1)|| \leqslant \varepsilon\}.$$

Then $P\{E_n(\varepsilon)\}\to 1$ as $n\to\infty$. On $E_n(\varepsilon)$, we have that, for each $t=1,\ldots,n-n'+1$, $|h_{n',t} - \hat{h}_{n',t}| \leqslant W_n(t;\varepsilon)$, where

$$W_n(t;\varepsilon) := (N')^{-1} \sum_{j=t}^{t+N'-1} \mathbf{1}(|\|m_{j,j+l'-1} - m_{t,t+n'-1}\|_\infty \sqrt{l'} - \|m_{t,t+n'-1} - m\|_\infty \sqrt{n'}| \leqslant \varepsilon).$$

By stationarity, we have that

$$\mathbb{E}|h_{n',t} - \hat{h}_{n',t}|\,\mathbf{1}\{E_n(\varepsilon)\} \leqslant (N')^{-1} \sum_{j=1}^{N'} P(|\|m_{j,j+l'-1} - m_{1,n'}\|_\infty \sqrt{l'} - \|m_{1,n'} - m\|_\infty \sqrt{n'}| \leqslant \varepsilon)$$
$$\to L(b,\varepsilon) := (1-b)^{-1}\int_0^{1-b} V_b(r,\varepsilon)\,\mathrm{d}r.$$

The above convergence follows from theorem 3 of Ferguson (1996) and the fact that $W_n(1;\varepsilon) \to_{\mathrm{D}} J(b,\varepsilon)$, where

$$J(b,\varepsilon) := (1-b)^{-1}\int_0^{1-b} \mathbf{1}\{|\sup_{s\in\mathbb{R}} |K(s,r+b) - K(s,r) - b\,K(s,1)|/\sqrt{b} - \sup_{s\in\mathbb{R}} |K(s,1)|| \leqslant \varepsilon\}\,\mathrm{d}r.$$

Again the continuous mapping theorem is invoked to derive the weak convergence of $W_n(1,\varepsilon)$ and, following the argument in the proof of theorem 1, its use can be justified under the assumption that $\tilde{V}_b(r,\varepsilon)=0$ for each $r\in[0,1-b]$ and $\varepsilon\geqslant 0$.

Next it is not difficult to see that $\lim_{\varepsilon\downarrow 0} L(b,\varepsilon)=0$ since $V_b(r,\varepsilon)\downarrow V_b(r,0)=0$ as $\varepsilon\downarrow 0$ for every $r\in[0,1-b]$. Thus $\sup_{t=1,\ldots,n-n'+1} \mathbb{E}|h_{n',t} - \hat{h}_{n',t}| \leqslant \mathbb{E}|h_{n',1} - \hat{h}_{n',1}|\,\mathbf{1}\{E_n(\varepsilon)\} + 2P\{E_n(\varepsilon)^c\} \leqslant \varepsilon$ for sufficiently large n. Furthermore,

$$\hat{J}_{n,n'}(x-\sqrt{\varepsilon}) - (n-n'+1)^{-1} \sum_{t=1}^{n-n'+1} \mathbf{1}(|\hat{h}_{n',t} - h_{n',t}| \geqslant \sqrt{\varepsilon}) \leqslant J_{n,n'}(x)$$
$$\leqslant \hat{J}_{n,n'}(x+\sqrt{\varepsilon}) + (n-n'+1)^{-1} \sum_{t=1}^{n-n'+1} \mathbf{1}(|\hat{h}_{n',t} - h_{n',t}| \geqslant \sqrt{\varepsilon}).$$

By the Markov inequality,

$$(n-n'+1)^{-1}\sum_{t=1}^{n-n'+1} P(|\hat{h}_{n',t}-h_{n',t}|\geqslant\sqrt{\varepsilon})\leqslant\sqrt{\varepsilon}.$$

Using the same argument as in the proof of theorem 3.2.1 of Politis *et al.* (1999a), we can show that $\hat{J}_{n,n'}(x)-P\{J(b)\leqslant x\}=o_p(1)$, which follows from the stationarity and strong mixing properties of X_t and the boundness of $\{\hat{h}_{n',t}\}_{t=1}^{n-n'+1}$. The conclusion then follows from an elementary argument.

References

Azaïs, J. M. and Wschebor, M. (2001) On the regularity of the distribution of the maximum of one-parameter Gaussian processes. *Probab. Theor. Reltd Flds*, **119**, 70–98.

Beran, R. (1987) Previoting to reduce level error of confidence sets. *Biometrika*, **74**, 457–468.

Beran, R. (1988) Prepivoting test statistics: a bootstrap view of asymptotic refinements. *J. Am. Statist. Ass.*, **83**, 687–697.

Berg, A., McMurry, T. and Politis, D. N. (2010) Subsampling p-values. *Statist. Probab. Lett.*, **80**, 1358–1364.

Berkes, I., Hörmann, S. and Schauer, J. (2009) Asymptotic results for the empirical process of stationary sequences. *Stoch. Processes Appl.*, **119**, 1298–1324.

Bickel, P. and Sakov, A. (2008) On the choice of *m* in the *m* out of *n* bootstrap and confidence bounds for extrema. *Statist. Sin.*, **18**, 967–985.

Billingsley, P. (1968) *Convergence of Probability Measures*. New York: Wiley.

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

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

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

Diebolt, J. and Posse, C. (1996) On the density of the maximum of smooth Gaussian processes. *Ann. Probab.*, **24**, 1104–1129.

Ferguson, T. S. (1996) *A Course in Large Sample Theory*. New York: Chapman and Hall.

Goncalves, S. and Vogelsang, T. J. (2011) Block bootstrap HAC robust tests: sophistication of the naive bootstrap. *Econmetr. Theor.*, **27**, 745–791.

Götze, F. and Račkauskas, A. (2001) Adaptive choice of bootstrap sample sizes. In *State of the Art in Probability and Statistics* (eds A. van der Vaart, M. de Gunst and C. Klaassen), pp. 286–309. Cambridge: Cambridge University Press.

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

Hall, P., Lahiri, S. N. and Jing, B.-Y. (1998) On the subsampling window method for long-range dependent data. *Statist. Sin.*, **8**, 1189–1204.

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

Hashimzade, N. and Vogelsang, T. J. (2008) Fixed-b asymptotic approximation of the sampling behavior of nonparametric spectral density estimators. *J. Time Ser. Anal.*, **29**, 142–162.

Jach, A., McElroy, A. and Politis, D. N. (2012) Subsampling inference for the mean of heavy-tailed long memory time series. *J. Time Ser. Anal.*, **33**, 96–111.

Jansson, M. (2004) The error rejection probability of simple autocorrelation robust test. *Econometrica*, **72**, 937–946.

Kiefer, N. M. and Vogelsang, T. J. (2005) A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econmetr. Theor.*, **21**, 1130–1164.

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

Lahiri, S. N. (2001) Effects of block lengths on the validity of block resampling methods. *Probab. Theor. Reltd Flds*, **121**, 73–97.

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

Lee, S. M. S. and Lai, P. Y. (2009) Double block bootstrap confidence intervals for dependent data. *Biometrika*, **96**, 427–443.

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

Loh, W.-Y. (1987) Calibrating confidence coefficients. *J. Am. Statist. Ass.*, **82**, 155–162.

Loh, W.-Y. (1991) Bootstrap calibration for confidence interval construction and selection. *Statist. Sin.*, **1**, 477–491.

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

Nordman, D. and Lahiri, S. N. (2005) Validity of sampling window method for linear long-range dependent processes. *Econmetr. Theor.*, **21**, 1087–1111.

Paparoditis, E. and Politis, D. N. (2001) Tapered block bootstrap. *Biometrika*, **88**, 1105–1119.
Paparoditis, E. and Politis, D. N. (2002) The tapered block bootstrap for general statistics from stationary sequences. *Econmetr. J.*, **5**, 131–148.
Politis, D. N. (2011) Higher-order accurate, positive semi-definite estimation of large-sample covariance and spectral density matrices. *Econmetr. Theor.*, **27**, 703–744.
Politis, D. N. and Romano, J. P. (1992) A circular block-resampling procedure for stationary data. In *Exploring the Limits of Bootstrap* (eds R. LePage and L. Billard), pp. 263–270. New York: Wiley.
Politis, D. N. and Romano, J. P. (1994) Large sample confidence regions based on subsamples under minimal assumptions. *Ann. Statist.*, **22**, 2031–2050.
Politis, D. N., Romano, J. P. and Wolf, M. (1999a) *Subsampling*. New York: Springer.
Politis, D. N., Romano, J. P. and Wolf, M. (1999b) Weak convergence of dependent empirical measures with application to subsampling in function spaces. *J. Statist. Planng Inf.*, **79**, 179–190.
Politis, D. N. and White, H. (2004) Automatic block-length selection for the dependent bootstrap. *Econmetr. Rev.*, **23**, 53–70.
Sayginsoy, O. and Vogelsang, T. J. (2011) Testing for a shift in trend at an unknown date: a fixed-b analysis of heteroskedasticity autocorrelation robust OLS based tests. *Econmetr. Theor.*, **27**, 992–1025.
Shao, X. (2010a) A self-normalized approach to confidence interval construction in time series. *J. R. Statist. Soc.* B, **72**, 343–366; correction, 695–696.
Shao, X. (2010b) Extended tapered block bootstrap. *Statist. Sin.*, **20**, 807–821.
Shao, X. (2010c) The dependent wild bootstrap. *J. Am. Statist. Ass.*, **105**, 218–235.
Shao, X. and Politis, D. (2011) Fixed-*b* subsampling and block bootstrap: improved confidence sets based on p-value calibration. *Technical Report*. Department of Statistics, University of Illinois at Urbana–Champaign, Champaign.
Sun, Y., Phillips, P. C. B. and Jin, S. (2008) Optimal bandwidth selection in heteroscedasticity-autocorrelation robust testing. *Econometrica*, **76**, 175–194.