Convergence rates of empirical block length selectors for block bootstrap

DANIEL J. NORDMAN[1] and SOUMENDRA N. LAHIRI[2]

[1]*Department of Statistics, Iowa State University, Ames, Iowa 50011, USA. E-mail: dnordman@iastate.edu*
[2]*Department of Statistics, Texas A&M University, College Station, Texas 77843, USA.*
E-mail: snlahiri@stat.tamu.edu

We investigate the accuracy of two general non-parametric methods for estimating optimal block lengths for block bootstraps with time series – the first proposed in the seminal paper of Hall, Horowitz and Jing (*Biometrika* **82** (1995) 561–574) and the second from Lahiri *et al.* (*Stat. Methodol.* **4** (2007) 292–321). The relative performances of these general methods have been unknown and, to provide a comparison, we focus on rates of convergence for these block length selectors for the moving block bootstrap (MBB) with variance estimation problems under the smooth function model. It is shown that, with suitable choice of tuning parameters, the optimal convergence rate of the first method is $O_p(n^{-1/6})$ where n denotes the sample size. The optimal convergence rate of the second method, with the same number of tuning parameters, is shown to be $O_p(n^{-2/7})$, suggesting that the second method may generally have better large-sample properties for block selection in block bootstrap applications beyond variance estimation. We also compare the two general methods with other plug-in methods specifically designed for block selection in variance estimation, where the best possible convergence rate is shown to be $O_p(n^{-1/3})$ and achieved by a method from Politis and White (*Econometric Rev.* **23** (2004) 53–70).

Keywords: jackknife-after-bootstrap; moving block bootstrap; optimal block size; plug-in methods; subsampling

1. Introduction

Performance of block bootstrap methods critically depends on the choice of block lengths. A common approach to the problem is to choose a block length that minimizes the Mean Squared Error (MSE) function of block bootstrap estimators as a function of the block length. For many important functionals, expansions for the MSE-optimal block lengths are known. If $\hat{\theta}_n$ denotes an estimator of a parameter of interest $\theta \in \mathbb{R}$ based on a stationary stretch $X_1, \ldots, X_n$, examples of relevant functionals φ_n of the distribution of $\hat{\theta}_n$ include the bias $\varphi_{1n} = \mathrm{E}(\hat{\theta}_n - \theta)$, variance $\varphi_{2n} = \mathrm{Var}(\hat{\theta}_n)$, and the distribution function $\varphi_{3n}(x_0) = P(\sqrt{n}(\hat{\theta}_n - \theta)/\tau_n \leq x_0)$ (i.e., given $x_0 \in \mathbb{R}$ and where τ_n^2 represents either the variance of $\sqrt{n}(\hat{\theta}_n - \theta)$ or an estimator of this, cf. [11]). If $\hat{\varphi}_n(\ell)$ denotes a block bootstrap estimator of φ_n based on block length ℓ, then as $n \to \infty$ the bias and variance of $\hat{\varphi}_n(\ell)$ often admit expansions of the form

$$n^{2a} \mathrm{Var}\big(\hat{\varphi}_n(\ell)\big) = V_0 \frac{\ell^r}{n}\big(1 + o(1)\big), \qquad n^a \mathrm{Bias}\big(\hat{\varphi}_n(\ell)\big) = -\frac{B_0}{\ell}\big(1 + o(1)\big) \qquad (1.1)$$

for some known constants $a, r > 0$ depending on φ_n (e.g., $a = r = 1$ for functionals $\varphi_n = \varphi_{1n}, \varphi_{2n}$, while $r = 2, a = 1/2$ for the distribution function $\varphi_n = \varphi_{3n}(x_0)$ when $|x_0| \neq 1$) and

lead to a large sample approximation of MSE-optimal block size given by

$$\ell_n^0 \equiv \ell_n^0(\varphi) = \mathcal{C}_0 n^{1/(r+2)}\big(1 + o(1)\big), \qquad \mathcal{C}_0 \equiv \left(\frac{2B_0^2}{rV_0}\right)^{1/(r+2)}, \tag{1.2}$$

involving population quantities $B_0 = B_0(\varphi_n)$, $V_0 = V_0(\varphi_n) \in \mathbb{R}$ that depend on the functional φ_n, the bootstrap method, and various parameters of the underlying process. For smooth function model statistics $\hat{\theta}_n$ (described below), these expansions (1.1) have been established for the moving block and non-overlapping block methods [5,7,8,11] and, in particular, are also known for the variance functional φ_{2n} with other block bootstraps, such as the circular block bootstrap [18] and stationary bootstrap [13,19]; see [10] and references therein. However, as the theoretical approximations (1.2) for the optimal block lengths typically depend on different unknown population parameters of the underlying process in an intricate manner, these are not directly usable in practice.

Different data-based methods for the selection of optimal block lengths have been proposed in the literature. One of the most popular general methods is proposed by Hall, Horowitz and Jing [5] (hereafter referred to as HHJ) which employs a subsampling method (cf. [19]) to construct an empirical version of the MSE function and minimizes this to produce an estimator of the optimal block length. We will refer to this approach as the HHJ method. A second general method for selecting the optimal block length is put forward by Lahiri *et al.* [11]. This method is based on the jackknife-after-bootstrap method of Efron [3] and its extension to block bootstrap by Lahiri [9]. For reasons explained in [11] (see also Section 2 below), we will refer to this method as the non-parametric plug-in method (or the NPPI method, in short). Both the HHJ and NPPI methods are called "general" because these can be used in the same manner across different functionals (e.g., bias, variance, distribution function, quantiles, etc.) to find the optimal block size for bootstrap estimation, *without* requiring exact analytical expressions for the corresponding optimal block length approximation (1.2) (i.e., without requiring explicit forms for quantities B_0, V_0). In particular, for a given functional, the HHJ method aims to directly estimate the constant $\mathcal{C}_0$ in the optimal block approximation (1.2) while the NPPI method separately and non-parametrically estimates the bias B_0 and variance V_0 quantities in (1.2) without structural knowledge of these. Our major objective here is to investigate the convergence rates of these two *general* methods. For instance, despite the popularity of the HHJ method, little is theoretically known about its properties for block selection or how this compares to the NPPI method. As a context to compare the methods, we focus on their performance for block selections in *variance* estimation problems with the block bootstrap. In the literature, a few other block length selection methods also exist. These are primarily plug-in estimators which *necessarily* require an explicit expression for the optimal block approximation (1.2) for each specific functional and for each block bootstrap method (i.e., requiring exact forms for B_0, V_0) and are not the focus of this paper. However, two popular plug-in methods for the variance functional in the latter category are given by Bühlmann and Künsch [2] and Politis and White [21] (and its corrected version Patton, Politis and White [16]). For completeness, we later compare the performance of the two general methods with these plug-in methods for block selection in variance estimation.

For concreteness, we shall restrict attention to the moving block bootstrap (MBB) method [7, 12], which was the original focus of the HHJ method [5] and the plug-in method of Bühlmann

and Künsch [2] and shares close large-sample connections to other block bootstrap methods (e.g., circular block bootstrap, non-overlapping block bootstrap, untapered version of the tapered block bootstrap) [8,13,15,21]. Further, we shall work under the *smooth function model* of Hall [4] (see Section 2.1 below) which provides a convenient theoretical framework but, at the same time, is general enough to cover many commonly used estimators in the time series context ([10]; Chapter 4). Accordingly, let $\hat{\theta}_n$ be an estimator of a parameter of interest θ under the smooth function model and suppose that the MBB is used for estimating $\sigma_n^2 \equiv n \operatorname{Var}(\hat{\theta}_n)$ or its limiting form

$$\sigma_\infty^2 \equiv \lim_{n \to \infty} n \operatorname{Var}(\hat{\theta}_n). \tag{1.3}$$

Let

$$\mathrm{MSE}_n(\ell) \equiv \mathrm{E}\big\{\hat{\sigma}_n^2(\ell) - \sigma_\infty^2\big\}^2 \tag{1.4}$$

denote the MSE of the MBB variance estimator $\hat{\sigma}_n^2(\ell)$ based on blocks of length ℓ and a sample of size n. (Defining the MSE with σ_n^2 or σ_∞^2 makes no difference in the following and, for clarity, it is helpful to fix a target σ_∞^2 in defining (1.4) throughout.)

The theoretical MSE-optimal block size is given by

$$\ell_n^{\mathrm{opt}} = \operatorname{argmin}\big\{\mathrm{MSE}_n(\ell) \colon \ell \in \mathcal{J}_n\big\}, \tag{1.5}$$

where $\mathcal{J}_n$ is a suitable set of block lengths including the optimal block length. As alluded to above (1.1), under some standard regularity conditions, it can be shown that

$$\mathrm{MSE}_n(\ell) \approx f_n(\ell) \equiv B_0^2 \ell^{-2} + V_0 n^{-1} \ell, \qquad \ell \in \mathcal{J}_n,$$

where B_0 and V_0 are population parameters arising, respectively, from the bias and variance of the MBB variance estimator $\hat{\sigma}_n^2(\ell)$. Let $\ell_n^0 \equiv \operatorname{argmin}\{f_n(\ell) \colon \ell > 0\} = \mathcal{C}_0 n^{1/3}$ denote the minimizer of the asymptotic approximation $f_n(\cdot)$ to the MSE function, where $\mathcal{C}_0 = [2B_0^2/V_0]^{1/3}$ (cf. (1.2)). As a first step towards investigating the accuracy of different empirical block rule selection methods, we consider the relative error of this theoretical approximation and show that

$$\frac{\ell_n^{\mathrm{opt}} - \ell_n^0}{\ell_n^0} = O\big(n^{-1/3}\big)$$

as $n \to \infty$. Thus, the true optimal block size and the optimal block size determined by the asymptotic approximation to the MSE curve of the block bootstrap estimator differ by a margin of $O(n^{-1/3})$ on the relative scale. In general, this rate cannot be improved further. As a result, for empirical block length selection rules involving estimation steps that target ℓ_n^0 (which all existing methods do), the upper bound on their accuracy for estimating the true optimal block length ℓ_n^{opt} is $O_p(n^{-1/3})$.

Next, we consider the convergence rates of the two general methods. Let $\hat{\ell}_{n,\mathrm{HHJ}}^{\mathrm{opt}}$ and $\hat{\ell}_{n,\mathrm{NPPI}}^{\mathrm{opt}}$, respectively, denote the estimators of the optimal block length based on the HHJ and NPPI methods. We show that under some mild conditions and with a suitable choice of the tuning parame-

ters,

$$\frac{\hat{\ell}^{\text{opt}}_{n,\text{HHJ}} - \ell^{\text{opt}}_n}{\ell^{\text{opt}}_n} = O_p\big(n^{-1/6}\big)$$

as $n \to \infty$. Thus, the (relative) rate of convergence of the HHJ estimator of the optimal block length is $O_p(n^{-1/6})$. The block length in block bootstrap methodology plays a role similar to a smoothing parameter in non-parametric functional estimation. It is well known (cf. [6]) that non-parametric data based rules for bandwidth estimation often have an "excruciatingly slow" (relative) rate of convergence (e.g., of the order of $O_p(n^{-1/10})$). The convergence rate of the HHJ method turns out to be relatively better. It is worth noting that the HHJ block estimator, based on the overlapping version of the subsampling method, has the *same* rate of convergence irrespective of the dependence structure of the underlying time series $\{X_t\}$. Additionally, in the process of determining this convergence rate, we also provide the theoretical guidance on optimally choosing two tuning parameters required in implementing the HHJ method, which has been an unresolved aspect of the method.

Next, we consider the NPPI method and compare its relative performance with the HHJ method. The rate of convergence of the NPPI method is determined by two factors, which arise from estimating the variance and the bias of a block bootstrap estimator (i.e., quantities V_0 and B_0 appearing in $\ell^0_n = C_0 n^{1/3}$, $C_0 = [2B_0^2/V_0]^{1/3}$). The factor due to the variance part is based on the (block) jackknife-after-bootstrap method [3,9], and it attains an optimal rate of $O_p(n^{-2/7})$, with a suitable choice of the tuning parameters. On the other hand, the second factor is determined by a non-standard bias estimator that turns out to be adaptive to the strength of dependence of $\{X_t\}$. Let $r(k)$ denote the autocovariance function of (a suitable linear function of) the X_t's. When $r(k) \sim Ck^{-a}$ as $k \to \infty$ for a suitably large $a > 1$, the rate of convergence of the second term can be as small as $O_p(n^{-1/2+\varepsilon})$, for a given $\varepsilon > 0$, with a suitable choice of the tuning parameters. Thus, combining the two, the optimal rate of convergence of the NPPI method becomes $O_p(n^{-2/7})$, which is better than optimal rate $O_p(n^{-1/6})$ for the HHJ method. For this to hold, the user needs to specify *two* tuning parameters, the same number as with the HHJ method. Also, the convergence rate $O_p(n^{-2/7})$ is interesting in the variance estimation problem because this matches the best rate obtained by the plug-in block selection method of Bühlmann and Künsch [2]. Their method is a four-step algorithm which uses lag weight estimators of the spectral density at zero and again requires explicit forms for quantities appearing in the bias and variance (e.g., B_0, V_0) of the MBB variance estimator. Hence, while the NPPI method for block selection applies more generally to other functionals, its convergent rate matches the optimal one for a plug-in method specifically tailored to the variance estimation problem. This provides some evidence supporting the use of the NPPI method in block selection with other functionals outside of variance estimation.

The rest of the paper is organized as follows. In Section 2, we briefly describe the smooth function model, the MBB and the empirical block length selectors proposed by HHJ [5] and Lahiri *et al.* [11]. In Section 3, we present the conditions and derive a general result on uniform approximation of the MSE of a block bootstrap estimator which may be of independent interest. We describe main results on the HHJ and the NPPI methods in Sections 4 and 5, respectively. In Section 6, we compare the general HJJ/NPPI methods with other plug-in block selection approaches for the MBB in the variance estimation problem. In particular, a plug-in method of

Politis and White [21] (see also [16]) is shown to achieve the best possible convergence rate for block selection with variance functionals. Section 7 sketches proofs of the main results, where full proofs are deferred to a supplementary material appendix [14].

2. Preliminaries

2.1. MBB variance estimator and optimal block length

Let $\mathcal{X}_n = (X_1, \ldots, X_n)$ be a stationary stretch of $\mathbb{R}^d$-valued random vectors with mean $\mathrm{E}X_t = \mu \in \mathbb{R}^d$. We shall consider the problem of estimating the variance of a statistic framed in the "smooth function" model [4]. Using some function $H : \mathbb{R}^d \to \mathbb{R}$ and the sample mean $\bar{X}_n = \sum_{i=1}^n X_i/n$, suppose that a statistic can be expressed as $\hat{\theta}_n = H(\bar{X}_n)$ for purposes of estimating a process parameter $\theta = H(\mu)$. The "smooth function" model covers a wide range of parameters and their estimators, including sample mean, sample autocovariances, Yule–Walker estimators, among others; see Chapter 4, [10] for more examples. Recall the target variance of interest is $\sigma_n^2 \equiv n \operatorname{Var}(\hat{\theta}_n)$ or its limit (1.3).

We next describe the MBB variance estimator. Let $\ell < n \in \mathbb{N}$ (set of positive integers) denote the block length and create overlapping length ℓ blocks from $\mathcal{X}_n$ as $\{\mathcal{X}_{i,\ell} : i = 1, \ldots, n - \ell + 1\}$, where $\mathcal{X}_{i,\ell} = (X_i, \ldots, X_{i+\ell-1})$ for any integer $i, \ell \geq 1$. We independently resample $\lfloor n/\ell \rfloor$ blocks by letting $I_1, \ldots, I_{\lfloor n/\ell \rfloor}$ denote i.i.d. random variables with a uniform distribution over block indices $\{1, \ldots, n - \ell + 1\}$ and then define a MBB sample $X_1^*, \ldots, X_{n_1}^*$ of size $n_1 = \ell \lfloor n/\ell \rfloor$ as $(\mathcal{X}_{I_1,\ell}, \ldots, \mathcal{X}_{I_{\lfloor n/\ell \rfloor},\ell})$, where $\lfloor x \rfloor$ denotes the integer part of a real number x. The MBB analog of $\hat{\theta}_n$ is given by $\hat{\theta}_n^* = H(\bar{X}_n^*)$ using the MBB sample mean $\bar{X}_n^* = \sum_{i=1}^{n_1} X_i^*/n_1$ and the MBB variance estimator is then defined as

$$\hat{\sigma}_n^2(\ell) \equiv n_1 \operatorname{Var}_*(\hat{\theta}_n^*),$$

where $\operatorname{Var}_*(\cdot)$ denotes the variance with respect to the bootstrap distribution conditional on the data $\mathcal{X}_n$.

For variance estimation, we briefly consolidate notation from Section 1 on optimal block lengths. The performance of the MBB again depends on the block choice ℓ. Under certain dependence conditions and block assumptions ($\ell^{-1} + \ell/n \to 0$), the asymptotic bias and variance of the MBB estimator are

$$\mathrm{E}\hat{\sigma}_n^2(\ell) - \sigma_\infty^2 = -\frac{B_0}{\ell}(1 + o(1)), \qquad \operatorname{Var}[\hat{\sigma}_n^2(\ell)] = V_0 \frac{\ell}{n}(1 + o(1)) \tag{2.1}$$

as $n \to \infty$, for some population parameters B_0, V_0 depending on the covariance structure of the underlying process (cf. [5,7] and Condition S of Section 3.1). Thus, the main component in MSE (1.4) of the MBB follows as

$$\operatorname{MSE}_n(\ell) \approx f_n(\ell) \equiv \frac{B_0^2}{\ell^2} + V_0 \frac{\ell}{n} \tag{2.2}$$

as $n \to \infty$. The minimizer of $f_n(\ell)$ is given by

$$\ell_n^0 \equiv \mathcal{C}_0 n^{1/3}, \tag{2.3}$$

where $\mathcal{C}_0 = [2B_0^2 / V_0]^{1/3}$. From (2.2) and (2.3), the optimal block minimizing $\mathrm{MSE}_n(\ell)$ behaves as $\ell_n^{\mathrm{opt}} \approx \ell_n^0 = \mathcal{C}_0 n^{1/3}$ in large samples [5,7,10]. As a result, to examine properties of the block length selection methods, we shall create a collection of block lengths $\mathcal{J}_n \equiv \{\ell \in \mathbb{N} : K^{-1} n^{1/3} \le \ell \le K n^{1/3}\}$, for a suitably large constant $K > 0$ such that $K^{-1} < \mathcal{C}_0 < K$, and formally define the optimal block size ℓ_n^{opt} as in (1.5).

2.2. The Hall–Horowitz–Jing (HHJ) block estimation method

The HHJ [5] method seeks to estimate the optimal block size ℓ_n^{opt} by minimizing an empirical version of the MSE (1.4) created by subsampling (data blocking). Let $m \equiv m_n \in \mathbb{N}$ denote a sequence satisfying $m^{-1} + m/n \to 0$ as $n \to \infty$, which serves to define the length of subsamples $\mathcal{X}_{i,m} = (X_i, \ldots, X_{i+m-1}), i = 1, \ldots, n - m + 1$. For each subsample, let $\hat{\sigma}_{i,m}^2(b)$ denote the MBB variance estimator resulting from resampling length b blocks from observations $\mathcal{X}_{i,m}$. For clarity, note that MBB block lengths on size m subsamples are denoted by "b," while "ℓ" denotes MBB block lengths applied to the original data $\mathcal{X}_n$. To approximate the error $\mathrm{MSE}_m(b) \equiv \mathrm{E}\{\hat{\sigma}_m^2(b) - \sigma_\infty^2\}^2$ in MBB variance estimation incurred by using length b blocks in samples of size m, we form a subsampling estimator

$$\widehat{\mathrm{MSE}}_m(b) = \frac{1}{n - m + 1} \sum_{i=1}^{n-m+1} \left[\hat{\sigma}_{i,m}^2(b) - \hat{\sigma}_n^2(\tilde{\ell}_n) \right]^2, \tag{2.4}$$

where the initializing MBB estimator $\hat{\sigma}_n^2(\tilde{\ell}_n)$ of σ_∞^2 is based on the entire sample $\mathcal{X}_n$ and on a plausible pilot block size $\tilde{\ell}_n$. By minimizing $\widehat{\mathrm{MSE}}_m(b)$ over $\mathcal{J}_m$, we formulate

$$\hat{b}_{m,\mathrm{HHJ}}^{\mathrm{opt}} = \mathrm{argmin}\{\widehat{\mathrm{MSE}}_m(b) : b \in \mathcal{J}_m\} \tag{2.5}$$

as an estimator of the theoretically optimal MBB block length b_m^{opt} for a size m sample, with

$$b_m^{\mathrm{opt}} = \mathrm{argmin}\{\mathrm{MSE}_m(b) : b \in \mathcal{J}_m\}. \tag{2.6}$$

Next, is a rescaling step that involves approximating true optimal block length ℓ_n^{opt} with the minimizer ℓ_n^0 of MSE-approximation (2.2). That is, as b_m^{opt} is the "size m sample version" of ℓ_n^{opt} in (1.5), one uses the large-sample block approximation $b_m^{\mathrm{opt}} \approx b_m^0 = \mathcal{C}_0 m^{1/3}$ and $\ell_n^{\mathrm{opt}} \approx \ell_n^0 = \mathcal{C}_0 n^{1/3}$ from (2.3) to re-scale $\hat{b}_{m,\mathrm{HHJ}}^{\mathrm{opt}}$ and subsequently define the HHJ estimator of ℓ_n^{opt} as

$$\hat{\ell}_{n,\mathrm{HHJ}}^{\mathrm{opt}} = (n/m)^{1/3} \hat{b}_{m,\mathrm{HHJ}}^{\mathrm{opt}}. \tag{2.7}$$

Hence, the HHJ method requires specifying both a subsample size m and a pilot MBB block size $\tilde{\ell}_n$, which impact the performance of the block estimator $\hat{\ell}_{n,\mathrm{HHJ}}^{\mathrm{opt}}$.

2.2.1. *An oracle-like subsampling MSE*

For purposes of comparison with the HHJ method, we also define a second subsampling MSE given as

$$\widehat{\mathrm{MSE}}_m^\infty(b) = \frac{1}{n-m+1} \sum_{i=1}^{n-m+1} \left[\hat{\sigma}_{i,m}^2(b) - \sigma_\infty^2\right]^2, \tag{2.8}$$

which resembles the empirical MSE (2.4) after replacing the variance estimator $\hat{\sigma}_n^2(\tilde{\ell}_n)$ with its target σ_∞^2 from (1.3). This subsampling MSE serves to remove one tuning parameter $\tilde{\ell}_n$ in the original HHJ method by unrealistically assuming σ_∞^2 is known. However, we may parallel the performance of the HHJ block estimators $\hat{b}_{m,\mathrm{HHJ}}^{\mathrm{opt}}$ and $\hat{\ell}_{n,\mathrm{HHJ}}^{\mathrm{opt}}$ to their oracle-like counterparts

$$\hat{b}_{m,\mathrm{HHJ}}^{\mathrm{opt},\infty} = \mathrm{argmin}\{\widehat{\mathrm{MSE}}_m^\infty(b)\colon b \in \mathcal{J}_m\} \tag{2.9}$$

based on (2.8) and the resulting estimator of the optimal block length ℓ_n^{opt} given by

$$\hat{\ell}_{n,\mathrm{HHJ}}^{\mathrm{opt},\infty} = (n/m)^{1/3}\hat{b}_{m,\mathrm{HHJ}}^{\mathrm{opt},\infty}. \tag{2.10}$$

Both $\hat{\ell}_{n,\mathrm{HHJ}}^{\mathrm{opt}}$ and $\hat{\ell}_{n,\mathrm{HHJ}}^{\mathrm{opt},\infty}$ estimate the same optimal block size ℓ_n^{opt}, but the estimator $\hat{\ell}_{n,\mathrm{HHJ}}^{\mathrm{opt},\infty}$ is based on an *unbiased* subsampling criterion through knowledge of σ_∞^2, that is, $\mathrm{E}[\widehat{\mathrm{MSE}}_m^\infty(b)] = \mathrm{MSE}_m(b)$ for all $b \in \mathcal{J}_m$.

2.3. The non-parametric plug-in (NPPI) method

The NPPI method is based on the non-parametric plug-in principle [11] which yields estimators of MSE optimal smoothing parameters in general non-parametric function estimation problems. Here we describe the method for estimating the optimal block length for the variance functional using the MBB. Like any plug-in method, the target quantity for the NPPI method is the minimizer ℓ_n^0 of the MSE-approximation $f_n(\ell)$ of (2.2), which again is of the form $\ell_n^0 = \mathcal{C}_0 n^{1/3}$ from (2.3) with population parameters B_0 and V_0 in $\mathcal{C}_0 = [2B_0^2/V_0]^{1/3}$ determined by the bias and variance expansion (2.1) of the MBB variance estimator. The NPPI method estimates the bias and the variance of the MBB estimator non-parametrically, and then estimates B_0 and V_0 by inverting (2.1). Specifically, the method constructs estimators $\widehat{\mathrm{BIAS}}$ and $\widehat{\mathrm{VAR}}$ satisfying

$$\frac{\widehat{\mathrm{VAR}}}{\mathrm{Var}(\hat{\sigma}_n^2(\ell_1))} \xrightarrow{p} 1, \qquad \frac{\widehat{\mathrm{BIAS}}}{\mathrm{Bias}(\hat{\sigma}_n^2(\ell_2))} \xrightarrow{p} 1 \qquad \text{as } n \to \infty$$

for some block lengths ℓ_1 and ℓ_2 and defines $\hat{V}_0 = [n\ell_1^{-1}]\widehat{\mathrm{VAR}}$ and $\hat{B}_0 = \ell_2\widehat{\mathrm{BIAS}}$. Then, the NPPI estimator of the optimal block length is given by

$$\hat{\ell}_{\mathrm{NPPI}}^0 = [2\hat{B}_0^2/\hat{V}_0]^{1/3}n^{1/3}. \tag{2.11}$$

The bias estimator for the NPPI method is

$$\widehat{\text{BIAS}} = 2\big[\hat{\sigma}_n^2(\ell_2) - \hat{\sigma}_n^2(2\ell_2)\big]$$

and the variance estimator is constructed using the jackknife-after-bootstrap (JAB) method [3,9], due to its computational advantages. For completeness, we next briefly describe the details of the JAB variance estimator.

Remark 1. Politis and Romano [20] considered an estimator related to $\widehat{\text{BIAS}}$ above for bias-correcting the Bartlett spectral estimator (e.g., at the zero frequency, this Bartlett estimator is asymptotically equivalent to $\hat{\sigma}_n^2(\ell_2)$ and their corrected estimator is equivalent to $2\hat{\sigma}_n^2(2\ell_2) - \hat{\sigma}_n^2(\ell_2)$). It is also important to re-iterate that, while the NPPI block estimator is based on general forms (cf. (1.1), (2.1)) for the asymptotic bias and variance of a bootstrap estimator, the HHJ block estimator requires only the optimal block order (cf. (1.2), (2.3)) for minimizing the asymptotic MSE of a bootstrap estimator; in this sense, the HHJ method requires less large-sample information and could potentially be more general. At the same time, as the MSE-optimal block order is typically derived from asymptotic bias/variance quantities, both NPPI and HHJ methods are generally intended to apply for block selection with the same problems, particularly under the smooth function model.

2.3.1. *The jackknife-after-bootstrap variance estimator*

The JAB method was initially proposed by [3] to assess accuracy of bootstrap estimators for independent data, and was extended to the dependent case by [9]. A key advantage of the JAB method is that it does *not* require a second level of resampling; the JAB method produces a variance estimate of a block bootstrap estimator by merely regrouping the resampled blocks used in computing the original block bootstrap estimator [9].

Suppose that the goal is to estimate the variance of an MBB estimator $\hat{\varphi}_n(\ell)$ based on blocks of length ℓ. (For notational simplicity here, consider $\ell = \ell_1$ and $\hat{\varphi}_n(\ell) = \hat{\sigma}_n^2(\ell)$.) Let $m \equiv m_n$ be an integer such that $m \to \infty$ and $m/n \to 0$ as $n \to \infty$. Here, m denotes the number of bootstrap blocks to be deleted for the JAB. Set $N = n - \ell + 1$, $M = N - m + 1$ and for $i = 1, \ldots, M$, let $I_i = \{1, \ldots, N\} \setminus \{i, \ldots, i + m - 1\}$. Also, let $\mathcal{X}_{i,\ell} = (X_i, \ldots, X_{i+\ell-1})$, $i = 1, \ldots, N$ be the MBB blocks of size ℓ. The first step of the JAB is to define a jackknife version $\hat{\varphi}_n^{(i)} \equiv \hat{\varphi}_n^{(i)}(\ell)$ of $\hat{\varphi}_n(\ell)$ for each $i \in \{1, \ldots, M\}$. Then, the *ith block-deleted jackknife point value* $\hat{\varphi}_n^{(i)}$ is obtained by resampling $\lfloor n/\ell \rfloor$ blocks randomly, with replacement from the reduced collection $\{\mathcal{X}_{i,\ell}: \ j \in I_i\}$ and then by computing the corresponding block bootstrap variance estimator using the resulting resample.

Then, the JAB estimator of the variance of $\hat{\varphi}_n \equiv \hat{\varphi}_n(\ell)$ is given by

$$\widehat{\text{VAR}}_{\text{JAB}}(\hat{\varphi}_n) = \frac{m}{(N-m)} \frac{1}{M} \sum_{i=1}^{M} \big(\tilde{\varphi}_n^{(i)} - \hat{\varphi}_n\big)^2, \tag{2.12}$$

where $\tilde{\varphi}_n^{(i)} = m^{-1}[N\hat{\varphi}_n - (N - m)\hat{\varphi}_n^{(i)}]$ is the *ith block-deleted jackknife pseudo-value* of $\hat{\varphi}_n$, $i = 1, \ldots, M$.

3. Results on uniform expansion of the MSE

3.1. Assumptions

To develop MSE and other probabilistic expansions, we require conditions on the dependence structure of the stationary $\mathbb{R}^d$-valued process $\{X_t\}_{t\in\mathbb{Z}}$ and the smooth function H, described below. Condition D prescribes differentiability assumptions on the smooth function H, Condition M_r describes mixing/moment assumptions as a function of positive integer r, and Condition S entails certain covariance sums are non-zero. In particular, the sums in Condition S define the constant $\mathcal{C}_0 = [2B_0^2/V_0]^{1/3}$ in the large-sample optimal block approximation $\ell_n^0 = \mathcal{C}_0 n^{1/3}$ from (2.3). For $\nu = (\nu_1, \ldots, \nu_d) \in (\mathbb{N} \cup \{0\})^d$, write $\|\nu\|_1 = \sum_{i=1}^d \nu_i$ in the following.

Condition D. *The function $H : \mathbb{R}^d \to \mathbb{R}$ is 3-times continuously differentiable and $\max\{|\partial^\nu H(x)/(\partial x_1 \cdots \partial x_d)|\colon \|\nu\|_1 = 3\} \leq C(1+\|x\|^{a_0})$, $x = (x_1, \ldots, x_d)' \in \mathbb{R}^d$ for some $C > 0$ and integer $a_0 \geq 0$.*

Condition M_r. *For some $\delta > 0$, $\mathrm{E}\|X_1\|^{2r+\delta} < \infty$ and $\sum_{k=1}^\infty k^{2r-1}\alpha(k)^{\delta/(2r+\delta)} < \infty$, where $\alpha(\cdot)$ denotes the strong mixing coefficient of the process $\{X_t\}_{t\in\mathbb{Z}}$.*

Condition S. *$B_0 \equiv \sum_{k=-\infty}^\infty |k| r(k) \neq 0$ and $V_0 \equiv (4/3)\sigma_\infty^4 > 0$ for $\sigma_\infty^2 = \sum_{k=-\infty}^\infty r(k)$ in (1.3), where $r(k) = \mathrm{Cov}(\nabla' X_0, \nabla' X_k)$, $k \in \mathbb{Z}$ and $\nabla = (\partial H(\mu)/\partial x_1, \ldots, \partial H(\mu)/\partial x_d)'$ is the vector of first order partial derivatives of H at $\mathrm{E}X_1 = \mu$.*

Mixing and moment assumptions as formulated in Condition M_r are standard in investigating block resampling methods (cf. [10], Chapter 5). Typical expansions of the MSE of the MBB variance estimator often require H to be 2-times differentiable in the smooth function model, whereas Condition D requires slightly more in order to determine a finer expansion of this MSE. The assumptions on the process quantities B_0, V_0 in Condition S are mild and standard for the block bootstrap [5,7,8,11,13,15,19]; in particular, the assumption on B_0 is needed to rule out i.i.d. processes.

3.2. Main results

Recalling the MSE-approximation $f_n(\ell) \equiv \ell^{-2}B_0^2 + n^{-1}\ell V_0$ for the MBB variance estimator from (2.2) (with constants B_0, V_0 as in Condition S above), Theorem 1 below provides a more refined expansion of this MSE over a collection of block lengths, $\mathcal{J}_n = \{\ell \in \mathbb{N}\colon K^{-1}n^{1/3} \leq \ell \leq Kn^{1/3}\}$ as in (1.5) (cf. Section 2.1), of optimal order.

Theorem 1. *Suppose that Conditions D, M_r with $r = 6 + 2a_0$, and Condition S hold, where a_0 is as specified by Condition D. Then, as $n \to \infty$,*

(i) *for $f_n(\cdot)$ defined in (2.2),*

$$\max_{\ell \in \mathcal{J}_n} \left| \mathrm{MSE}_n(\ell) - \frac{2B_0\sigma_\infty^2}{n} - f_n(\ell) \right| = O(n^{-4/3}).$$

(ii) $|\ell_n^{\text{opt}} - \ell_n^0|/\ell_n^0 = O(n^{-1/3})$, *for* $\ell_n^0 \equiv \text{argmin}_{y>0} f_n(y) = C_0 n^{1/3}$ *from* (2.3).

Theorem 1(i) gives a close bound $O(n^{-4/3})$ how the MSE-approximation $f_n(\ell)$ matches the curve $\text{MSE}_n(\ell) - n^{-1}2B_0\sigma_\infty^2$ (*not* quite $\text{MSE}_n(\ell)$ but both having the same minimizer), uniformly in $\ell \in \mathcal{J}_n$. For comparison, note $f_n(\ell)$, $\ell \in \mathcal{J}_n$, has exact order $O(n^{-2/3})$. In trying to resolve ℓ_n^{opt}, we then have a general bound on the differences $n^{2/3}\{\text{MSE}_n(\ell) - \text{MSE}_n(\ell_n^{\text{opt}}) - [f_n(\ell) - f_n(\ell_n^{\text{opt}})]\} = O(n^{-2/3})$ between the two curves. One implication, stated in Theorem 1(ii), is that $O(n^{-1/3})$ becomes the general order on the discrepancy between the minimizer ℓ_n^{opt} of $\text{MSE}_n(\cdot)$ and the minimizer ℓ_n^0 of $f_n(\cdot)$. *Theorem 1 bounds cannot be generally improved by further expanding* $\text{MSE}_n(\ell)$ (i.e., under additional smoothness assumptions on H) and, in fact in Theorem 1(ii), ℓ_n^{opt} is necessarily an integer while ℓ_n^0 need not be.

4. Results on the HHJ method

To state the main result, recall $\hat{\ell}_{n,\text{HHJ}}^{\text{opt}}$ denotes the HHJ block estimator (2.7), depending on a pilot block $\tilde{\ell}_n$ and subsample size m, and that $\hat{\ell}_{n,\text{HHJ}}^{\text{opt},\infty}$ from (2.10) denotes an oracle-like version of $\hat{\ell}_{n,\text{HHJ}}^{\text{opt}}$ that requires m but not $\tilde{\ell}_n$.

Theorem 2. *Suppose that Conditions D, M_r with $r = 14 + 4a_0$, and Condition S hold, with a_0 as specified by Condition D. Assume that $m^{-1} + m/n \to 0$ as $n \to \infty$ with $m^{5/3}/n = O(1)$.*

(i) *Then, as $n \to \infty$,*

$$\left| \frac{\hat{\ell}_{n,\text{HHJ}}^{\text{opt},\infty} - \ell_n^{\text{opt}}}{\ell_n^{\text{opt}}} \right| = O_p\big(\max\{m^{-1/3}, m^{-1/12}(m/n)^{1/4}, (m/n)^{1/3}\}\big).$$

(ii) *If additionally $\tilde{\ell}_n^{-1} + \tilde{\ell}_n^2/n \to 0$ and $m/\tilde{\ell}_n^2 + m^2/n = O(1)$, then*

$$\left| \frac{\hat{\ell}_{n,\text{HHJ}}^{\text{opt}} - \ell_n^{\text{opt}}}{\ell_n^{\text{opt}}} \right| = O_p\left(\max\left\{m^{-1/3}, \frac{m^{1/3}}{\tilde{\ell}_n}, \frac{m^{1/6}}{n^{1/4}}, \frac{m^{1/3}}{(\tilde{\ell}_n n)^{1/4}}, \frac{\tilde{\ell}_n^{1/4}}{n^{1/4}}, \frac{m^{1/3}\tilde{\ell}_n^{1/2}}{n^{1/2}} \right\}\right).$$

Remark 2. Theorem 2 also holds if, on the left-hand sides above, we replace $(\hat{\ell}_{n,\text{HHJ}}^{\text{opt},\infty} - \ell_n^{\text{opt}})/\ell_n^{\text{opt}}$ and $(\hat{\ell}_{n,\text{HHJ}}^{\text{opt}} - \ell_n^{\text{opt}})/\ell_n^{\text{opt}}$ with their subsample counterparts $(\hat{b}_{m,\text{HHJ}}^{\text{opt},\infty} - b_m^{\text{opt}})/b_m^{\text{opt}}$ and $(\hat{b}_{m,\text{HHJ}}^{\text{opt}} - b_m^{\text{opt}})/b_m^{\text{opt}}$. This result helps to reinforce the notion that the quality of block estimation at the subsample level determines the performance of HHJ method.

Theorem 2(i) indicates how the subsample size m affects the convergence rate of the oracle-type block estimate. It follows from Theorem 1(i) that, with oracle knowledge of σ_∞^2, the best possible (fastest) rate of convergence for $(\hat{\ell}_{n,\text{HHJ}}^{\text{opt},\infty} - \ell_n^{\text{opt}})/\ell_n^{\text{opt}}$ is $O_p(n^{-1/6})$ achieved when the subsample size $m \propto n^{1/2}$. The choice $m \propto n^{1/2}$ balances the sizes of all three terms in the bound from Theorem 2(i). Remark 3 below provides some explanation of the probabilistic bounds in Theorem 2(i).

In Theorem 2(ii), we impose some additional block growth conditions on the pilot block $\tilde{\ell}_n$ and subsample size m in the HHJ method, which are mild and help to concisely express the order of the main components contributing to the error rate. While the combined effects of the tuning parameters are complicated and difficult to characterize in Theorem 2(ii), a block $\tilde{\ell}_n \propto n^{1/3}$ of MSE-optimal order for the pilot MBB variance estimator $\hat{\sigma}_n^2(\tilde{\ell}_n)$ in the HHJ method is an intuitive starting point. And with this choice, it follows that $m \propto n^{1/2}$ is then optimal for minimizing the convergence rate of the HHJ block estimator, which becomes $O_p(n^{-1/6})$. In fact, the selection $m \propto n^{1/2}$, $\tilde{\ell}_n \propto n^{1/3}$ is overall optimal and simultaneously balances the order $O_p(n^{-1/6})$ of *all six* error terms in Theorem 1(ii). So surprisingly, the HHJ block estimator $\hat{\ell}_{n,\text{HHJ}}^{\text{opt}}$ achieves the best convergence rate that one could hope for by matching the optimal rate of the oracle block estimator $\hat{\ell}_{n,\text{HHJ}}^{\text{opt},\infty}$. We summarize our findings on tuning parameters in Corollary 1.

Corollary 1. *Under the assumptions of Theorem 2, a subsample size $m \propto n^{1/2}$ and pilot block $\tilde{\ell}_n \propto n^{1/3}$ yield optimal convergence rates*

$$\left| \frac{\hat{\ell}_{n,\text{HHJ}}^{\text{opt},\infty} - \ell_n^{\text{opt}}}{\ell_n^{\text{opt}}} \right| = O_p(n^{-1/6}), \qquad \left| \frac{\hat{\ell}_{n,\text{HHJ}}^{\text{opt}} - \ell_n^{\text{opt}}}{\ell_n^{\text{opt}}} \right| = O_p(n^{-1/6})$$

as $n \to \infty$, for the HHJ block estimator $\hat{\ell}_{n,\text{HHJ}}^{\text{opt}}$ and its oracle $\hat{\ell}_{n,\text{HHJ}}^{\text{opt},\infty}$ version.

An interpretation of Corollary 1 is that, at optimal tuning parameters, random fluctuations in the HHJ block estimator $|\hat{\ell}_{n,\text{HHJ}}^{\text{opt}} - \ell_n^{\text{opt}}|$ are of the order $\sqrt{\ell_n^{\text{opt}}}$. This behavior interestingly resembles that of some other kernel bandwidth estimators based empirical MSE criteria (cf. [22]), though $\widehat{\text{MSE}}_m(\cdot)$ does not take its arguments from a continuum of real-values.

Remark 3. We provide a brief explanation of the probabilistic bounds in Theorem 2, and focus mainly on the behavior of oracle block estimator $\hat{b}_{m,\text{HHJ}}^{\text{opt},\infty}$ from Section 2.2.1 at the subsample level; more rigorous details are given in Section 7 and the supplementary material [14]. Recall the block estimator $\hat{b}_{m,\text{HHJ}}^{\text{opt},\infty}$ minimizes the $\widehat{\text{MSE}}_m^{\infty}(b)$ from (2.8), while b_m^{opt} from (2.6) minimizes $\text{MSE}_m(b) \approx f_m(b) \equiv b^{-1}B_0^2 + bm^{-1}V_0$ (the subsample version of (2.2)). In part, the bound $O(m^{-1/3})$ in Theorem 2(i) is due to smoothness issues with $\text{MSE}_m(b)$ and its discrepancy from $f_m(b)$ (cf. Theorem 1). The other bounds in Theorem 2 arise from the size of

$$\Delta_m^{\infty}(b) = \left\{ \widehat{\text{MSE}}_m^{\infty}(b_m^{\text{opt}}) - \text{E}\big[\widehat{\text{MSE}}_m^{\infty}(b_m^{\text{opt}})\big] \right\} - \left\{ \widehat{\text{MSE}}_m^{\infty}(b) - \text{E}\big[\widehat{\text{MSE}}_m^{\infty}(b)\big] \right\}, \qquad (4.1)$$

$b \in \mathcal{J}_m$, where $\text{E}[\widehat{\text{MSE}}_m^{\infty}(b)] = \text{MSE}_m(b)$; this quantity measures the discrepancy between two differenced curves (which should ideally match at $b = \hat{b}_{m,\text{HHJ}}^{\text{opt},\infty}$), where differences in $\text{MSE}_m(b)$ serve to identify b_m^{opt} and similar differences in $\widehat{\text{MSE}}_m^{\infty}(b)$ identify $\hat{b}_{m,\text{HHJ}}^{\text{opt},\infty}$. It can be shown that, for any $a_n \to 0$,

$$\max_{b \in \mathcal{J}_m: \, |b - b_m^{\text{opt}}| \le a_n m^{1/3}} a_n^{-1/2} m^{2/3} (n/m)^{1/2} \big| \Delta_m^{\infty}(b) \big|$$

remains stochastically bounded on shrinking neighborhoods of block lengths around b_m^{opt}, while at the same time $\hat{b}_{m,\text{HHJ}}^{\text{opt},\infty}/b_m^{\text{opt}} \xrightarrow{P} 1$ (i.e., $\hat{b}_{m,\text{HHJ}}^{\text{opt},\infty}$ is consistent for $b_m^{\text{opt}} \approx b_m^0 = \mathcal{C}_0 m^{1/3}$); see the auxiliary result, Theorem 6, of Section 7. This allows other order bounds on $(\hat{b}_{m,\text{HHJ}}^{\text{opt},\infty} - b_m^{\text{opt}})/b_m^{\text{opt}}$ to be determined by recursively "caging" $\hat{b}_{m,\text{HHJ}}^{\text{opt},\infty}$ in decreasing neighborhoods around b_m^{opt} with high probability. The probabilistic bounds in Theorem 2(ii) are partly due to error contributions from the MBB variance estimator $\hat{\sigma}_n^2(\tilde{\ell}_n)$ used through $\widehat{\text{MSE}}_m(b)$ in (2.4) to estimate $\text{MSE}_m(b)$ at the subsample level.

5. Results on the NPPI method

Next, we consider the convergence rates of the optimal block length selector based on the NPPI method. Recall that $r(k) = \text{Cov}(Y_1, Y_{k+1})$, $k \geq 1$, where $Y_i = \nabla' X_i$, $i \geq 1$.

Theorem 3. *Suppose that Conditions D, M_r with $r = 7 + 2a_0$, and Condition S hold, with a_0 as specified by Condition D. Assume that $\ell_2 n^{-1/3} + \ell_1^{-1} + \ell_1 m^{-1} + m/n \to 0$ as $n \to \infty$. Then, as $n \to \infty$,*

$$
\begin{aligned}
&\left|\hat{\ell}_{n,\text{NPPI}}^{\text{opt}} - \ell_n^{\text{opt}}\right|/\ell_n^{\text{opt}} \\
&= O_p\left([m/n]^{1/2} + [\ell_1/m] + \ell_1^{-2}\right) + O_p\left(\ell_2 \sum_{k=\ell_2}^{2\ell_2-1} |r(k)| + n^{-1/2}\ell_2^{3/2}\right).
\end{aligned}
\tag{5.1}
$$

As the NPPI method targets the block approximation $\ell_0 = [2B_0^2/V_0]^{1/3} n^{1/3}$ (2.3), the first of the two terms on the right side of (5.1) is from the estimation of V_0 and the second is from the estimation of B_0. For the first term, with any given choice of m, the optimal choice of ℓ_1 satisfies $\ell_1/m \propto \ell_1^{-2}$, that is, $\ell_1 \propto m^{1/3}$. For this choice of ℓ_1, the optimal choice of m is determined by the relation $[m/n]^{1/2} \propto m^{1/3}/m$, that is, $m \propto n^{3/7}$. Thus, the optimal rate of the first term is $O_p(n^{-2/7})$ with $m \propto n^{3/7}$ and $\ell_1 \propto n^{1/7}$.

To determine the optimal order of the second term, first note that the pilot block size ℓ_2 is only required to satisfy the constraints stated in Theorem 3. In particular, ℓ_2 is not required to go to ∞ with the sample size. From (5.1), it is also evident that the optimal choice of ℓ_2 (to minimize the order of the second term alone) depends on the rate of decay of the autocovariance function $r(\cdot)$. Since $r(k) \leq Ck^{-a-1}$ for some $a \geq 12$ (implied by Condition M_r with $r = 7 + 2a_0$), the second term can always be made to match the optimal order of the first term, that is, $O_p(n^{-2/7})$, by choosing $\ell_2 = O(n^{1/7})$ (note $n^{-1/2}\ell_2^{3/2} = n^{-2/7}$ in (5.1) when $\ell_2 \propto n^{1/7}$). However, for processes with an exponentially decaying $r(k)$, a choice of $\ell_2 \propto \log n$ optimizes the second term, with the attained rate of $O_p(n^{-1/2}[\log n]^{3/2})$, while for an m_0-dependent sequence $\{X_t\}$ with a fixed $m_0 \geq 1$, a choice of $\ell_2 = m_0 + 1$ makes the second term $O_p(n^{-1/2})$. But, in the end, the error rate $O_p(n^{-2/7})$ of first term dominates the second in (5.1).

Remark 4. In Lahiri *et al.* [11], the NPPI plug-in estimator was defined with a common choice $\ell_1 = \ell_2$. In this case, under the conditions of Theorem 3, the optimal order of the common block

size is determined by $O_p([m/n]^{1/2} + [\ell_1/m] + \ell_1^{-2} + n^{-1/2}\ell_1^{3/2})$. For a fixed ℓ_1, the first two factors are optimized for

$$m \propto n^{1/3}\ell_1^{2/3}.$$

Interestingly, this order of m was also suggested by [11], purely on the basis of some heuristic arguments. For this choice of m, one may choose $\ell_1 \propto n^{1/7}$ to optimize the rate of convergence of the NPPI method, yielding the same optimal rate $O_p(n^{-2/7})$ possible with three tuning parameters in the NPPI method. This supports the suggestion of Lahiri *et al.* [11] of a common choice $\ell_1 = \ell_2$ and, with the same number of tuning parameters, the NPPI block selector has a better optimal rate than $O_p(n^{-1/6})$ for the HHJ method.

We summarize our findings on the NPPI method in Corollary 2.

Corollary 2. *Under the assumptions of Theorem 3, a JAB block deletion size $m \propto n^{3/7}$ and tuning block lengths $\ell_1 \propto n^{1/7}$, $\ell_2 = O(n^{1/7})$ as $n \to \infty$ yield an optimal convergence rate for the NPPI block estimator $\hat{\ell}_{n,\mathrm{NPPI}}^{\mathrm{opt}}$ as*

$$\left| \frac{\hat{\ell}_{n,\mathrm{NPPI}}^{\mathrm{opt}} - \ell_n^{\mathrm{opt}}}{\ell_n^{\mathrm{opt}}} \right| = O_p(n^{-2/7}).$$

In particular, choosing $m \propto n^{3/7}$, $\ell_1 = \ell_2 \propto n^{1/7}$ achieves this optimal rate.

6. Comparison with plug-in methods and concluding remarks

In the problem choosing an appropriate block length for implementing block bootstraps in time series, the HHJ and NPPI methods represent the two existing *general* block selection methods in the literature. However, because convergence rates of these block estimators have been unknown, our goal here was to provide some comparison of their relative performances, considering block estimation for MBB variance estimation in particular. Both methods are again "general" in the sense that one could consider block ℓ estimation for a block bootstrap version $\hat{\varphi}_n(\ell)$ of a general functional φ_n (e.g., bias, variance, distribution function, quantiles, etc. as in Section 1) of the sampling distribution of a time series estimator $\hat{\theta}_n$, by replacing the MBB variance functional $\hat{\varphi}_n(\ell) = \hat{\sigma}_n^2(\ell) \equiv \ell\lfloor n/\ell \rfloor \mathrm{Var}_*(\hat{\theta}_n^*)$ with $\hat{\varphi}_n(\ell)$ in the mechanics of the HHJ and NPPI methods described in Sections 2.2–2.3. Both methods aim to estimate MSE-optimal block length through its large sample approximation $\ell_n^0 = \mathcal{C}_0 n^{1/(r+2)}, \mathcal{C}_0 = [2B_0^2/(rV_0)]^{1/(r+2)}$ in (1.2) and neither method requires explicit forms for population quantities $B_0 \equiv B_0(\varphi_n)$, $V_0 \equiv V_0(\varphi_n)$ (arising, resp., from the bias and variance of a bootstrap functional $\hat{\varphi}_n(\ell)$ in (1.1)) which can depend on the functional φ_n and unknown process parameters in a complex way. The HHJ approach estimates the constant $\mathcal{C}_0$ in ℓ_n^0 directly through a subsampling technique, while the NPPI method non-parametrically estimates both B_0 and V_0 in $\mathcal{C}_0$. Intuitively, because the general NPPI approach separately targets the bootstrap bias/variance B_0, V_0 contributions to $\mathcal{C}_0$, one might anticipate this approach to exhibit better convergence rates in block estimation compared to HHJ. In considering block estimation for MBB variance estimation with time series, we have shown that this is

indeed the case. For the variance problem, NPPI achieves a better rate $O_p(n^{-2/7})$ than the HHJ method $O_p(n^{-1/6})$ when both methods use two tuning parameters. While considering the MBB among possible block bootstrap approaches, the same convergence rates and optimal tuning parameter selections should also hold for other block bootstraps, such as the non-overlapping block bootstrap [7], the circular block bootstrap [18] and the stationary bootstrap [13,19] (though the tapered block bootstrap [15] requires a different treatment as the bias expansion in (1.2) or (2.1) needs to be replaced by a smaller bias term $\ell^{-2}B_0$ in variance estimation). And though we have focused on variance estimation, we suspect that the NPPI method retains similar large-sample superiority over the HHJ method for block selection in other inference problems.

As mentioned in the Introduction, in the particular setting of block bootstrap variance estimation, other plug-in methods for block selection exist such as the proposals of Bühlmann and Künsch (BK) [2] and Politis and White (PW) [21] (see also Patton, Politis and White [16]). These use explicit expressions for the bias B_0 and variance components V_0 of the MBB variance estimator from (2.1) appearing the approximation $\ell_n^0 = [2B_0^2/V_0]n^{1/3}$ of the MSE-optimal block length ℓ_n^{opt} (1.5), given in this case by

$$B_0 = \sum_{k=-\infty}^{\infty} |k|r(k), \qquad V_0 = \frac{4}{3}\left(\sum_{k=-\infty}^{\infty} r(k)\right)^4 \tag{6.1}$$

for $r(k) = \text{Cov}(\nabla'X_1, \nabla'X_{1+k})$, $k \geq 1$; see Condition S, Section 3.1. The BK and PW approaches estimate the covariance sums (6.1) with spectral lag window estimators which are then plugged into the approximation ℓ_n^0 (2.3) to estimate ℓ_n^{opt}. The BK method is based on an iterative plug-in algorithm from Bühlmann [1] for estimating the optimal bandwidth for lag window estimators of the spectral density at zero, which has equivalences to block length selection for the MBB variance estimator. If $\hat{\ell}_{n,\text{BK}}^{\text{opt}}$ denotes the resulting block estimator, results in Bühlmann and Künsch [2] show that

$$\frac{\hat{\ell}_{n,\text{BK}}^{\text{opt}} - \ell_n^{\text{opt}}}{\ell_n^{\text{opt}}} = O_p(n^{-2/7})$$

for MBB variance estimation of smooth function model statistics. As mentioned earlier, interestingly, the NPPI block selection method obtains the exact *same* optimal rate of convergence without using the structural knowledge in (6.1). The plug-in estimator $\hat{\ell}_{n,\text{PW}}^{\text{opt}}$ of Politis and White [21] is formulated by using a "flat-top" lag-window $\lambda(t) = \mathbb{I}(t \in [0, 1/2]) + 2(1 - |t|)\mathbb{I}(t \in (1/2, 1])$, $t \in [0, 1]$, where $\mathbb{I}(\cdot)$ denotes the indicator function; see [20]. Their method was originally studied for block bootstrap variance estimation of time series sample means. Here we describe an extension of the methodology for smooth function model statistics $\hat{\theta}_n = H(\bar{X}_n)$. The corresponding two unknown covariance sums (6.1) in ℓ_n^0 are estimated, respectively, with

$$\sum_{k=-2M}^{2M} \lambda\{k/(2M)\}|k|\hat{\nabla}\hat{r}(k)\hat{\nabla}', \qquad \sum_{k=-2M}^{2M} \lambda\{k/(2M)\}\hat{\nabla}\hat{r}(k)\hat{\nabla}', \tag{6.2}$$

where $\hat{r}(k) = n^{-1} \sum_{i=1}^{n-|k|} (X_i - \bar{X}_n)(X_{i+|k|} - \bar{X}_n)'$, $\hat{\nabla} = \partial H(\bar{X}_n)/\partial x$, and M is a positive integer bandwidth. In which case, we may state a result on the convergence rate of the generalized PW block estimator.

Theorem 4. *Under the assumptions of Theorem* 1, *if* $M \propto n^\tau$ *as* $n \to \infty$ *for some* $10^{-1} \leq \tau \leq 3^{-1}$, *Politis–White block estimator* $\hat{\ell}_{n,\mathrm{PW}}^{\mathrm{opt}}$ *satisfies*

$$\left| \frac{\hat{\ell}_{n,\mathrm{PW}}^{\mathrm{opt}} - \ell_n^{\mathrm{opt}}}{\ell_n^{\mathrm{opt}}} \right| = O_p(n^{-1/3}).$$

This also holds for other rules for selecting M under Theorem 3.3 *conditions of* [21].

Remark 5. It should be noted that the rate in Theorem 4 differs from results in Politis and White [21] who considered a different problem in block estimation. Namely, they considered convergence rates between $\ell_n^0 = [2B_0^2/V_0]n^{1/3}$ and its plug-in counterpart $\hat{\ell}_{n,\mathrm{PW}}^{\mathrm{opt}} = [2\hat{B}_0^2/\hat{V}_0]n^{1/3}$, where ℓ_n^0 again represents the large-sample approximation (2.3) of the MSE-optimal block length ℓ_n^{opt} from (1.5). They showed that, depending on the underlying process dependence (cf. their Theorem 3.3), the bandwidth M can be adaptively chosen so that $|\hat{\ell}_{n,\mathrm{PW}}^{\mathrm{opt}} - \ell_n^0|/\ell_n^0$ may exhibit a convergence rate as high as $O_p(n^{-1/2})$; see also Politis [17] for a related discussion of rate adaptivity and empirical rules for selecting M. In these cases, there is still a bound $O(n^{-1/3})$ on the relative closeness of ℓ_n^0 and ℓ_n^{opt} from Theorem 1. Additionally, while the PW and NPPI methods both involve plug-in estimation, the NPPI approach does not require or use an explicit form for B_0, V_0 in the variance problem (6.1), and the discussion of Section 5 indicates that this method can adaptively estimate B_0 (with similar rates as high as $O_p(n^{-1/2})$) but does not adaptively estimate V_0. That is, the JAB (i.e., block jackknife) variance estimator for V_0 is not rate adaptive in the NPPI method, but the PW flat-top kernel approach is. These differences explain the superior performance of the PW method compared to NPPI for block estimation in the variance estimation problem.

Table 1 provides a final summary of the convergence rates of both general and (6.1)-based plug-in methods for block selection with the MBB variance estimator. The PW plug-in estimator attains the highest convergence rate $O_p(n^{-1/3})$ possible under Theorem 1 for any estimator of MSE-optimal block length ℓ_n^{opt} which is based on its asymptotic approximation ℓ_n^0 (2.3). That is,

Table 1. Optimal convergence rate $|\hat{\ell}_n^{\mathrm{opt}} - \ell_n^{\mathrm{opt}}|/\ell_n^{\mathrm{opt}}$ for block estimators $\hat{\ell}_n^{\mathrm{opt}}$ of the MSE-optimal block length ℓ_n^{opt} (1.5) for MBB variance estimation, based on the approximation ℓ_n^0 (2.3)

	Methods				
	General		Form (6.1)-based plug-in		
	HHJ	NPPI	Bühlmann–Künsch (BK)	Politis–White (PW)	Best possible
Rate	$O_p(n^{-1/6})$	$O_p(n^{-2/7})$	$O_p(n^{-2/7})$	$O_p(n^{-1/3})$	$O_p(n^{-1/3})$

the plug-in method of Politis and White [21] has the *best* large-sample properties of any existing method for block selection in the variance estimation problem with mean-like or smooth function model statistics. Of course, this advantage comes at the price in that the PW method is designed for variance estimation (i.e., the forms (6.1) in this problem) and is therefore "non-general" or not directly usable for block selection in other block bootstrap applications. In particular, for other inference problems (e.g., distribution or quantile estimation), the forms of B_0, V_0 in the large-sample block formulas (1.2) can become complicated, depending additionally sums of higher order process cumulants in a more complex fashion than the variance estimation problem. In these cases, where appropriate block selections for the block bootstrap are still needed, the general HHJ and NPPI block estimation methods have their greatest appeal, and the convergence rate results in variance estimation suggest that the NPPI may have better performance than HHJ more generally.

7. Additional results and proofs

Theorem 5 below gives a bias and variance decomposition for the MBB variance estimator $\hat{\sigma}_m^2(b)$, uniformly in $b \in \mathcal{J}_m$, which is used to establish Theorem 1. The proof of Theorem 5 appears in the supplementary material [14].

Theorem 5. *Under the assumptions of Theorem 1, as $m \to \infty$,*

$$\text{(i)} \qquad \max_{b \in \mathcal{J}_m} \left| \left[\mathrm{E}\hat{\sigma}_m^2(b) - \sigma_\infty^2 \right] + \left(\frac{B_0}{b} + \frac{b}{m}\sigma_\infty^2 \right) \right| = O(m^{-1}),$$

$$\text{(ii)} \qquad \max_{b \in \mathcal{J}_m} \left| \mathrm{Var}\left[\hat{\sigma}_m^2(b) \right] - V_0 \frac{b}{m} \right| = O(m^{-4/3}).$$

Proof of Theorem 1. Part (i) follows directly from Theorem 5. Part (ii) follows by expanding $0 \le n^{2/3}[\mathrm{MSE}_n(\lfloor \ell_n^0 \rfloor) - \mathrm{MSE}_n(\ell_n^{\mathrm{opt}})]$ with Theorem 1(i), implying $0 \le n^{2/3}[f_n(\ell_n^{\mathrm{opt}}) - f_n(\ell_n^0)] \le Cn^{-2/3}$. Then, a second order Taylor expansion of $f_n(\cdot)$ around $\ell_n^0 = C_0 n^{1/3}$ gives the result (as $df_n(\ell_n^0)/dy = 0$). $\qquad \square$

Theorem 6 next establishes the consistency of the HHJ block estimator (and its oracle-version) at both sample and subsample levels, and provides tightness results for developing rates for the HHJ block estimator (cf. Theorem 2); its proof is given in the supplementary material [14]. To state the result, recall the difference $\Delta_m^\infty(b)$, $b \in \mathcal{J}_m$, between empirical and true MSE curves from (4.1) and define $\Delta_m(b)$ by replacing $\widehat{\mathrm{MSE}}_m^\infty(\cdot)$ from (2.8) with $\widehat{\mathrm{MSE}}_m(\cdot)$ from (2.4) in (4.1) (i.e., HHJ method uses $\widehat{\mathrm{MSE}}_m$). Then,

$$\Delta_m(b) = \Delta_m^\infty(b) + \Omega_{1,m}(b) + \Omega_{2,m}(b), \qquad \Omega_{1,m}(b) \equiv \Omega_{3,m}(b) - \mathrm{E}\left[\Omega_{3,m}(b)\right] \qquad (7.1)$$

holds for $b \in \mathcal{J}_m$, where $\Omega_{2,m}(b) \equiv 2[\hat{\sigma}_n^2(\tilde{\ell}_n) - \mathrm{E}\hat{\sigma}_n^2(\tilde{\ell}_n)]\mathrm{E}[\hat{\sigma}_{1,m}^2(b_m^{\mathrm{opt}}) - \hat{\sigma}_{1,m}^2(b)]$ and

$$\Omega_{3,m}(b) \equiv 2\frac{\hat{\sigma}_n^2(\tilde{\ell}_n) - \sigma_\infty^2}{n-m+1} \sum_{i=1}^{n-m+1} \left\{ \left[\hat{\sigma}_{i,m}^2(b) - \hat{\sigma}_{i,m}^2(b_m^{\mathrm{opt}})\right] - \mathrm{E}\left[\hat{\sigma}_{i,m}^2(b) - \hat{\sigma}_{i,m}^2(b_m^{\mathrm{opt}})\right] \right\}.$$

Given any $C > 0$, define a block set $\mathcal{J}_m^{\mathrm{opt}}(C) = \{b \in \mathcal{J}_m \colon |b_m^{\mathrm{opt}} - b| \leq Cm^{1/3}\}$.

Theorem 6. *Suppose that Conditions D, M_r with $r = 14 + 4a_0$, and Condition S hold, with a_0 as specified by Condition D. Assume that $m^{-1} + m/n \to 0$ with $m^{5/3}/n = O(1)$ as $n \to \infty$ and that $\tilde{\ell}_n$ in the HJJ method satisfies $\tilde{\ell}_n^{-1} + \tilde{\ell}_n^2/n \to 0$ and $m(\tilde{\ell}_n^{-2} + n^{-1}\tilde{\ell}_n) = O(1)$. Let $\Lambda_m(b)$, $b \in \mathcal{J}_m$ denote either $\Delta_m^\infty(b)$ or $\Delta_m^\infty(b) + \Omega_{1,m}(b)$. Then,*

(i) *there exists an integer $N_0 \geq 1$ and constant $A > 0$ such that*

$$P\left(a_n^{-1/2}m^{2/3}\left(\frac{n}{m}\right)^{1/2} \max_{b \in \mathcal{J}_m^{\mathrm{opt}}(a_n)} \left|\Lambda_m(b)\right| > \lambda\right) \leq \frac{A}{\lambda},$$

$$P\left(a_n^{-1}m^{2/3}\left(\frac{m^{1/3}}{\tilde{\ell}_n}\frac{n}{m}\right)^{1/2} \max_{b \in \mathcal{J}_m^{\mathrm{opt}}(a_n)} \left|\Omega_{2,m}(b)\right| > \lambda\right) \leq \frac{A}{\lambda},$$

holds for any $\lambda > 0$, any $n \geq N_0$ and any positive $a_n > 0$.

(ii) $\hat{b}_{m,\mathrm{HHJ}}^{\mathrm{opt}}/b_m^{\mathrm{opt}} \xrightarrow{P} 1$ *and* $\hat{b}_{m,\mathrm{HHJ}}^{\mathrm{opt},\infty}/b_m^{\mathrm{opt}} \xrightarrow{P} 1$ *as* $n \to \infty$.

(iii) $\hat{\ell}_{n,\mathrm{HHJ}}^{\mathrm{opt}}/\ell_n^{\mathrm{opt}} \xrightarrow{P} 1$ *and* $\hat{\ell}_{n,\mathrm{HHJ}}^{\mathrm{opt},\infty}/\ell_n^{\mathrm{opt}} \xrightarrow{P} 1$ *as* $n \to \infty$.

Proof of Theorem 2. We establish Theorem 2(i) here and defer the proof of Theorem 2(ii) to the supplementary material [14]. For the minimizer b_m^{opt} of $\mathrm{MSE}_m(\cdot)$ from (2.6) and the minimizer $b_m^0 = \mathcal{C}_0 m^{1/3}$ of $f_m(y)$, $y > 0$ (i.e., subsample version of (2.3) solving $d[f_m(b_m^0)]/dy = 0$), Theorem 1(ii) gives $m^{-1/3}|b_m^0 - b_m^{\mathrm{opt}}| = O(m^{-1/3})$, $m^{2/3}|f_m(b_m^0) - f_m(b_m^{\mathrm{opt}})| = O(m^{-2/3})$ so that

$$0 \leq m^{2/3}\left[f_m\left(\hat{b}_{m,\mathrm{HHJ}}^{\mathrm{opt},\infty}\right) - f_m\left(b_m^0\right)\right] \leq Cm^{-2/3} + m^{2/3}\left[\mathrm{MSE}_m\left(\hat{b}_{m,\mathrm{HHJ}}^{\mathrm{opt},\infty}\right) - \mathrm{MSE}_m\left(b_m^{\mathrm{opt}}\right)\right], \quad (7.2)$$

by Theorem 1(i), for a constant $C > 0$ independent of m. Applying a Taylor expansion of $f_m(\hat{b}_{m,\mathrm{HHJ}}^{\mathrm{opt},\infty})$ around b_m^0 and Theorem 1(ii), there exists a constant $C_0 > 0$ for which

$$m^{-2/3}\left(\hat{b}_{m,\mathrm{HHJ}}^{\mathrm{opt},\infty} - b_m^{\mathrm{opt}}\right)^2 \leq C_0 \max\left\{m^{-2/3}, m^{2/3}\left[\mathrm{MSE}_m\left(\hat{b}_{m,\mathrm{HHJ}}^{\mathrm{opt},\infty}\right) - \mathrm{MSE}_m\left(b_m^{\mathrm{opt}}\right)\right]\right\},$$

whenever $|\hat{b}_{m,\mathrm{HHJ}}^{\mathrm{opt},\infty}/b_m^0 - 1| < 1/2$. Also, by definition we have

$$0 \leq m^{2/3}\left[\mathrm{MSE}_m\left(\hat{b}_{m,\mathrm{HHJ}}^{\mathrm{opt},\infty}\right) - \mathrm{MSE}_m\left(b_m^{\mathrm{opt}}\right)\right] \leq m^{2/3}\Delta_m^\infty\left(\hat{b}_{m,\mathrm{HHJ}}^{\mathrm{opt},\infty}\right) \quad (7.3)$$

for $\Delta_m^\infty(\cdot)$ defined in (4.1) where $\mathrm{E}[\widehat{\mathrm{MSE}}_m^\infty(b)] = \mathrm{MSE}_m(b)$, $b \in \mathcal{J}_m$ so that

$$m^{-2/3}\left(\hat{b}_{m,\mathrm{HHJ}}^{\mathrm{opt},\infty} - b_m^{\mathrm{opt}}\right)^2 \leq C_1 \max\left\{m^{-2/3}, m^{2/3}\Delta_m^\infty\left(\hat{b}_{m,\mathrm{HHJ}}^{\mathrm{opt},\infty}\right)\right\}, \quad (7.4)$$

must hold for any $C_1 > C_0$ whenever $|\hat{b}_{m,\mathrm{HHJ}}^{\mathrm{opt},\infty}/b_m^0 - 1| < 1/2$; since this last event has arbitrarily large probability by Theorem 6(ii), we will always assume (7.4) to hold without loss of generality along with $\mathcal{J}_m^{\mathrm{opt}}(C_0) = \mathcal{J}_m$, defining a block set $\mathcal{J}_m^{\mathrm{opt}}(C) = \{b \in \mathcal{J}_m: |b_m^{\mathrm{opt}} - b| \le Cm^{1/3}\}$ for any $C > 0$.

We next formulate a series of recursive events to coerce $\hat{b}_{m,\mathrm{HHJ}}^{\mathrm{opt},\infty}$ into shrinking neighborhoods around b_m^{opt} with high probability. Fix $C_1 > C_0$ and define $a_{0,n} = C_1$, $L \equiv \lceil \log\log n \rceil > 1$, and

$$a_{i,n}^2 \equiv C_1^2 \max\left\{ m^{-2/3}, 2^{\sum_{k=0}^{i-1}[(L-i+1)+k]4^{-k}} \left(\frac{m}{n}\right)^{2^{-1}\sum_{k=0}^{i-1}4^{-k}} \right\}, \qquad i \ge 1.$$

Define an integer $J = \min\{i = 1, \ldots, L+1: a_{i,n}^2 = C_1^2 m^{-2/3}\}$ and let $J = L+1$ if this integer set is empty. For $i = 0, \ldots, J-1$, let A_i be the event

$$\max_{b \in \mathcal{J}_m^{\mathrm{opt}}(a_{i,n})} m^{2/3}\left|\Delta_m^\infty(b)\right| \le a_{i,n}^{1/2}(m/n)^{1/2}\lambda_i, \qquad \lambda_i \equiv C_1^{1/2}2^{L-i}$$

and let B_i, $i \ge 1$, be the event

$$m^{-2/3}\left(\hat{b}_{m,\mathrm{HHJ}}^{\mathrm{opt},\infty} - b_m^{\mathrm{opt}}\right)^2 \le a_{i,n}^2.$$

Since $\mathcal{J}_m^{\mathrm{opt}}(a_{0,n}) = \mathcal{J}_m$, event A_0 implies B_1 by (7.4). Also, for $J > 1$, if $A_i \cap B_i$ holds for some $i = 1, \ldots, J-1$, then so must B_{i+1} by (7.4), which in turn implies $\hat{b}_{m,\mathrm{HHJ}}^{\mathrm{opt},\infty}$ in the block neighborhood $\mathcal{J}_m^{\mathrm{opt}}(a_{i+1,n})$ for event A_{i+1}. Suppose now that $A_J \cap B_J$ holds for an event A_J defined as

$$\max_{b \in \mathcal{J}_m^{\mathrm{opt}}(a_{J,n})} m^{2/3}\left|\Delta_m^\infty(b)\right| \le a_{n,J}^{1/2}(m/n)^{1/2}C_1^{1/2};$$

the complement $(A_J \cap B_J)^c$ has probability bounded by $\sum_{i=0}^J P(A_i^c) \le AC_1^{-1/2}(1 + \sum_{k=0}^L 2^{-k}) \le 3AC_1^{-1/2}$ by Theorem 6(i), which can be made arbitrarily small by large C_1. When $A_J \cap B_J$ holds, then by construction (7.4) further implies that either

$$m^{-2/3}\left(\hat{b}_{m,\mathrm{HHJ}}^{\mathrm{opt},\infty} - b_m^{\mathrm{opt}}\right)^2 \le C_1^2 \max\left\{m^{-2/3}, m^{-1/6}(m/n)^{1/2}\right\}$$

if $a_{J,n}^2 = C_1^2 m^{-2/3}$, or the remaining possibility is $a_{J,n} \ne C_1^2 m^{-2/3}$ in event A_J and $J = L+1$ so that

$$m^{-2/3}\left(\hat{b}_{m,\mathrm{HHJ}}^{\mathrm{opt},\infty} - b_m^{\mathrm{opt}}\right)^2 \le C_1^2 \max\left\{ m^{-2/3}, 2^{\sum_{k=0}^L k4^{-k}} \left(\frac{m}{n}\right)^{2^{-1}\sum_{k=0}^L 4^{-k}} \right\}$$

$$\le 2^{3/2}C_1^2 \max\left\{ m^{-2/3}, \left(\frac{m}{n}\right)^{2/3} \right\}$$

using that $(m/n)^{2^{-1}\sum_{k=0}^L 4^{-k}-2/3} \le 2$ and $2^{\sum_{k=1}^L k4^{-k}} \le \sqrt{2}$ for $L = \lceil \log\log n \rceil$. Hence,

$$m^{-2/3}\left(\hat{b}_{m,\mathrm{HHJ}}^{\mathrm{opt},\infty} - b_m^{\mathrm{opt}}\right)^2 \le 4C_1^2 \max\left\{m^{-2/3}, m^{-1/6}(m/n)^{1/2}, (m/n)^{2/3}\right\}$$

holds with arbitrarily high probability (large C_1). Because $|b_m^{\text{opt}} - b_m^0|m^{-1/3} = O(m^{-1/3})$ and $|\ell_n^{\text{opt}} - \ell_n^0|n^{-1/3} = O(n^{-1/3})$ by Theorem 1(ii), where $\ell_n^0 \equiv C_0 n^{1/3} = b_m^0 (n/m)^{1/3}$ and $\hat{\ell}_{n,\text{HHJ}}^{\text{opt},\infty} = \hat{b}_{m,\text{HHJ}}^{\text{opt},\infty} (n/m)^{1/3}$ is formed by rescaling (2.10), Theorem 2(i) follows. $\qquad\square$

Proof of Theorem 3. We sketch the proof, providing more technical detail in the supplementary material [14]. Considering $\hat{V}_0$ and letting $p = n/\ell_1$, it can be shown that

$$\left|\widehat{\text{VAR}} - \text{Var}(\hat{\sigma}_n^2(\ell_1))\right| = O_p\left(p^{-1}[[m/n]^{1/2} + \ell/m]\right). \tag{7.5}$$

Next using the arguments in the proof of Theorem 5(ii), one can show that

$$\text{Var}(\hat{\sigma}_n^2(\ell_1)) = \frac{V_0}{p} + O\left(\ell_1^{-1}n^{-1}\right). \tag{7.6}$$

Hence, by (7.5) and (7.6), it follows that

$$\begin{aligned}
|\hat{V}_0 - V_0| &= \left|n\ell_1^{-1}\widehat{\text{VAR}} - V_0\right| \\
&\leq \left|n\ell_1^{-1}\left(\widehat{\text{VAR}} - \text{Var}(\hat{\sigma}_n^2(\ell_1))\right)\right| + \left|n\ell_1^{-1}\text{Var}(\hat{\sigma}_n^2(\ell_1)) - V_0\right| \\
&= O_p\left([m/n]^{1/2} + [\ell_1/m] + \ell_1^{-2}\right).
\end{aligned}$$

Next consider $\hat{B}_0$. Using arguments in the proof of Theorem 5, one can show that

$$\text{E}\hat{\sigma}_n^2(k) = k\text{E}[\bar{Y}_k]^2 + O\left(n^{-1}k\right), \qquad \text{Var}(\hat{\sigma}_n^2(k)) = O\left(n^{-1}k\right)$$

for $k = \ell_2, 2\ell_2$, where $\bar{Y}_k = k^{-1}\sum_{i=1}^k Y_i$ and $Y_i = \nabla' X_i$, $i \geq 1$. Hence, it follows that

$$\begin{aligned}
\hat{B}_0 &= 2\ell_2\left(\ell_2\text{E}[\bar{Y}_{\ell_2}]^2 - 2\ell_2\text{E}[\bar{Y}_{2\ell_2}]^2\right) + O_p\left(n^{-1/2}\ell_2^{3/2}\right) \\
&= B_0 + O\left(\sum_{k=\ell_2}^{2\ell_2-1} k|r(k)| + \ell_2\sum_{k=\ell_2}^{2\ell_2-1}|r(k)|\right) + O_p\left(n^{-1/2}\ell_2^{3/2}\right).
\end{aligned}$$

Combining the bounds on $\hat{V}_0$ and $\hat{B}_0$, the theorem follows. $\qquad\square$

Proof of Theorem 4. From the assumed conditions, $\hat{\nabla} \equiv \partial H(\bar{X}_n)/\partial x = \nabla + O_p(n^{-1/2})$ holds for $\nabla \equiv \partial H(\mu)/\partial x$. By (6.1)–(6.2), the PW block estimator for MBB variance estimation can be written as $n^{-1/3}\hat{\ell}_{n,\text{PW}}^{\text{opt}} = n^{-1/3}\tilde{\ell}_{n,\text{PW}} + O_p(n^{-1/2})$ for

$$\tilde{\ell}_{n,\text{PW}} = (3/2)^{1/3}\left(\sum_{k=-2M}^{2M} \lambda\{k/(2M)\}|k|\hat{r}_Y(k)\right)^{2/3}\left(\sum_{k=-2M}^{2M} \lambda\{k/(2M)\}\hat{r}_Y(k)\right)^{-4/3} n^{1/3}$$

with $\hat{r}_Y(k) = n^{-1}\sum_{i=1}^{n-|k|}(Y_i - \bar{Y}_n)^2$ with $Y_i = \nabla' X_i$, $i \geq 1$. By Theorem 3.3(i) of [21] with $M \propto n^\tau$ for some $10^{-1} \leq \tau \leq 3^{-1}$ and the assumed mixing conditions here, $|\tilde{\ell}_{n,\text{PW}}^{\text{opt}} - \ell_n^0|/\ell_0 =$

$O_p(n^{-(1-\tau)/2}) = O_p(n^{-1/3})$ follows for $\ell_n^0 = [2B_0^2/V_0]^{1/3}n^{1/3}$ with B_0, V_0 in (6.1). Hence, by Theorem 1(ii) and $\ell_n^0/\ell_n^{\mathrm{opt}} \to 1$,

$$\frac{|\hat{\ell}_{n,\mathrm{PW}}^{\mathrm{opt}} - \ell_n^{\mathrm{opt}}|}{\ell_n^{\mathrm{opt}}} \leq \frac{|\hat{\ell}_{n,\mathrm{PW}}^{\mathrm{opt}} - \tilde{\ell}_{n,\mathrm{PW}}|}{\ell_n^{\mathrm{opt}}} + \frac{|\tilde{\ell}_{n,\mathrm{PW}} - \ell_n^0|}{\ell_n^{\mathrm{opt}}} + \frac{|\ell_n^0 - \ell_n^{\mathrm{opt}}|}{\ell_n^{\mathrm{opt}}}$$

$$= O_p(n^{-1/2}) + O_p(n^{-1/3}) + O(n^{-1/3}) = O_p(n^{-1/3}). \qquad \square$$

Acknowledgments

Supplementary Material

Proofs of main results for empirical block length selectors (DOI: 10.3150/13-BEJ511SUPP; .pdf). A supplement [14] provides more detailed proofs of the main results (Theorems 2–3) about the convergence rates for the HHJ/NPPI block selection methods from Sections 4–5, as well as proofs for the auxiliary results (Theorems 5–6) of Section 7.

References

[1] Bühlmann, P. (1996). Locally adaptive lag-window spectral estimation. *J. Time Series Anal*. **17** 247–270. MR1391180

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

[3] Efron, B. (1992). Jackknife-after-bootstrap standard errors and influence functions. *J. Roy. Statist. Soc. Ser. B* **54** 83–127. MR1157715

[4] Hall, P. (1992). *The Bootstrap and Edgeworth Expansion. Springer Series in Statistics*. New York: Springer. MR1145237

[5] Hall, P., Horowitz, J.L. and Jing, B.Y. (1995). On blocking rules for the bootstrap with dependent data. *Biometrika* **82** 561–574. MR1366282

[6] Hall, P. and Marron, J.S. (1987). On the amount of noise inherent in bandwidth selection for a kernel density estimator. *Ann. Statist*. **15** 163–181. MR0885730

[7] Künsch, H.R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist*. **17** 1217–1241. MR1015147

[8] Lahiri, S.N. (1999). Theoretical comparisons of block bootstrap methods. *Ann. Statist*. **27** 386–404. MR1701117

[9] Lahiri, S.N. (2002). On the jackknife-after-bootstrap method for dependent data and its consistency properties. *Econometric Theory* **18** 79–98. MR1885351

[10] Lahiri, S.N. (2003). *Resampling Methods for Dependent Data. Springer Series in Statistics*. New York: Springer. MR2001447

[11] Lahiri, S.N., Furukawa, K. and Lee, Y.D. (2007). A nonparametric plug-in rule for selecting optimal block lengths for block bootstrap methods. *Stat. Methodol*. **4** 292–321. MR2380557

[12] Liu, R.Y. and Singh, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring the Limits of Bootstrap* (*East Lansing*, *MI*, 1990) (R. Lepage, and L. Billard, eds.). *Wiley Ser. Probab. Math. Statist. Probab. Math. Statist*. 225–248. New York: Wiley. MR1197787

[13] Nordman, D.J. (2009). A note on the stationary bootstrap's variance. *Ann. Statist*. **37** 359–370. MR2488355

[14] Nordman, D.J. and Lahiri, S.N. (2014). Supplement to "Convergence rates of empirical block length selectors for block bootstrap." DOI:10.3150/13-BEJ511SUPP.

[15] Paparoditis, E. and Politis, D.N. (2001). Tapered block bootstrap. *Biometrika* **88** 1105–1119. MR1872222

[16] Patton, A., Politis, D.N. and White, H. (2009). Correction to "Automatic block-length selection for the dependent bootstrap" by D. Politis and H. White. *Econometric Rev*. **28** 372–375. MR2532278

[17] Politis, D.N. (2003). Adaptive bandwidth choice. *J. Nonparametr. Stat*. **15** 517–533. MR2017485

[18] Politis, D.N. and Romano, J.P. (1992). A circular block-resampling procedure for stationary data. In *Exploring the Limits of Bootstrap* (*East Lansing*, *MI*, 1990). *Wiley Ser. Probab. Math. Statist. Probab. Math. Statist*. 263–270. New York: Wiley. MR1197789

[19] Politis, D.N. and Romano, J.P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *Ann. Statist*. **22** 2031–2050. MR1329181

[20] Politis, D.N. and Romano, J.P. (1995). Bias-corrected nonparametric spectral estimation. *J. Time Series Anal*. **16** 67–103. MR1323618

[21] Politis, D.N. and White, H. (2004). Automatic block-length selection for the dependent bootstrap. *Econometric Rev*. **23** 53–70. MR2041534

[22] Rice, J. (1984). Bandwidth choice for nonparametric regression. *Ann. Statist*. **12** 1215–1230. MR0760684