Department of Economics

Sieve Bootstrap for Strongly Dependent Stationary Processes

George Kapetanios and Zacharias Psaradakis

Working Paper No. 552 January 2006 ISSN 1473-0278

Queen Mary
University of London

SIEVE BOOTSTRAP FOR STRONGLY DEPENDENT STATIONARY PROCESSES

George Kapetanios*
Queen Mary, University of London

Zacharias Psaradakis†
Birkbeck College, University of London

January 2006

Abstract

This paper studies the properties of the sieve bootstrap for a class of linear processes which exhibit strong dependence. The sieve bootstrap scheme is based on residual resampling from autoregressive approximations the order of which increases slowly with the sample size. The first-order asymptotic validity of the sieve bootstrap is established in the case of the sample mean and sample autocovariances. The finite-sample properties of the method are also investigated by means of Monte Carlo experiments.

Keywords: Autoregressive approximation; linear process; strong dependence; sieve bootstrap; stationary process.
JEL Codes: C10, C22, C50

*Department of Economics, Queen Mary, University of London, Mile End Road, London E1 4NS, U.K.; e-mail: `g.kapetanios@qmul.ac.uk`.

†School of Economics, Mathematics and Statistics, Birkbeck College, University of London, Malet Street, London WC1E 7HX, U.K.; e-mail: `z.psaradakis@bbk.ac.uk`.

1 Introduction

The autoregressive sieve bootstrap was proposed by Kreiss (1992) and Bühlmann (1997) as a means of obtaining consistent estimates of the variances and distributions of statistics associated with dependent data. The idea is to approximate the (possibly infinite-dimensional) data-generating mechanism by an autoregressive model the order of which increases to infinity simultaneously but sufficiently slowly with the sample size; this autoregressive approximation may then be used to resample residuals and generate bootstrap replicates of the data. The properties of the sieve bootstrap scheme have been rigorously investigated by Kreiss (1992), Paparoditis (1996), Bühlmann (1997), Bickel and Bühlmann (1999), and Choi and Hall (2000), among others, who established its asymptotic validity for a variety of statistics under the assumption that the data come from an infinite-order autoregressive process.

A common assumption in all the papers cited above is that the coefficients in the moving-average representation of the stochastic process of interest are absolutely summable, or satisfy even stronger summability conditions. Such assumptions ensure that the process exhibits weak or short-range dependence in the sense of having autocorrelations which decay fast enough to be absolutely summable.

The aim of the present paper is to extend this literature by exploring the behaviour of the sieve bootstrap for stochastic processes that exhibit strong or long-range dependence. The characterising feature of such processes is that their autocorrelations tend to zero hyperbolically in the lag parameter and hence are not absolutely summable (for a general survey of the properties of strongly dependent processes see Beran (1994)). Stochastic processes that exhibit strong dependence have been found to be useful for modelling real-world time series occurring in many fields, including economics, hydrology, geophysics and telecommunications. We show that, under appropriate regularity conditions, the sieve bootstrap provides an asymptotically valid approximation to the distribution of the sample mean and sample autocovariances for a large class of strongly dependent linear processes with square-summable coefficients declining at slow hyperbolic rates.

It is worth noting briefly here existing work on bootstrap procedures for strongly dependent data. For processes obtained through instantaneous transformations of strongly dependent stationary Gaussian sequences, Lahiri (1993) demonstrated that the moving-blocks bootstrap provides an asymptotically valid approximation to the distribution of the sample mean only when latter is asymptotically normal. For the same class of processes, Hall, Jing, and Lahiri (1998) showed that consistent estimation of the distribution of the sample mean can be achieved by means of block-subsampling, while Nordman and Lahiri (2005) established validity of the method for stationary linear processes. Andrews and Lieberman (2002) presented results which show that the parametric bootstrap can provide improvements upon the asymptotic approximation to distributions of covariance parameter estimates for strongly dependent stationary Gaussian processes. Hidalgo (2003) proposed a semiparametric bootstrap procedure for the coefficients of regression equations involving strongly dependent stationary processes, which is based on resampling in the frequency domain.

Our paper contributes to this literature by providing a procedure for inference on (functions of) the sample mean and sample autocovariances of strongly dependent processes based on a semiparametric time-domain resampling scheme. Our results imply that the sieve bootstrap can be applied for inference on stationary processes without imposing weak dependence conditions. Although the method relies on the assumption that the data come from a linear process which admits an autoregressive representation, existing results on the closure of the sets of infinite-order moving-average and autoregressive processes, as discussed in the work of Bickel and Bühlmann (1996) and Bickel and Bühlmann (1997), suggest that this requirement may not be too onerous.

The remainder of the paper is organised as follows. Section 2 introduces notation and the class of stochastic processes under consideration. The definition of the sieve bootstrap scheme is given in Section 3, where some of the probabilistic properties of the sieve bootstrap are also established. Sections 4 and 5 discuss the asymptotic validity of the sieve bootstrap in approximating the distribution of the sample mean and sample autocovariances, respectively. Section 6 presents the results of a simulation study of the small-sample properties of the sieve bootstrap. Section 7 contains some final remarks.

2 Notation and Assumptions

Let $\{y_t, t \in \mathbb{Z}\}$ be a real-valued stochastic process satisfying the equation[1]

$$y_t - \mu = (1 - L)^{-d} u_t, \qquad t \in \mathbb{Z}, \tag{1}$$

for some constants $\mu \in \mathbb{R}$ and $d \in (0, \frac{1}{2})$. Here, L denotes the lag operator ($Ly_t = y_{t-1}$) and $\{u_t, t \in \mathbb{Z}\}$ is a purely non-deterministic process satisfying

$$u_t = \pi(L)\varepsilon_t, \qquad t \in \mathbb{Z}, \tag{2}$$

[1] In the sequel, $\mathbb{C}$ and $\mathbb{R}$ denote, respectively, the set of complex numbers and the real line, $\mathbb{N} = \{1, 2, \ldots\}$, $\mathbb{Z}^+ = \{0\} \cup \mathbb{N}$, and $\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$.

with

$$\pi(z) = \sum_{j=1}^{\infty} \pi_j z^j, \qquad z \in \mathbb{C},$$

and $\pi_0 = 1$. The following assumptions about $\{\varepsilon_t, t \in \mathbb{Z}\}$ and $\{\pi_j, j \in \mathbb{Z}^+\}$ will be maintained throughout the paper.

(A1) $\{\varepsilon_t, t \in \mathbb{Z}\}$ is a sequence of independent and identically distributed (i.i.d.) real-valued random variables such that $E(\varepsilon_0) = 0$ and $E(|\varepsilon_0|^4) < \infty$.

(A2) $\{\pi_j, j \in \mathbb{Z}^+\}$ is a sequence of real numbers such that $\sum_{j=0}^{\infty} |\pi_j| < \infty$ and $\sum_{j=0}^{\infty} \pi_j z^j \neq 0$ for all $z \in \mathbb{C}$ with $|z| \leq 1$.

As usual, for any $d \notin \{0, -1, -2, \ldots\}$, the operator $(1 - L)^{-d}$ in (1) is defined by using the series expansion of $(1 - z)^{-d}$ ($|z| < 1$), i.e.,

$$(1 - L)^{-d} = 1 + \sum_{j=1}^{\infty} \frac{\Gamma(j + d)}{\Gamma(d)\Gamma(j + 1)} L^j,$$

where $\Gamma(\cdot)$ denotes the gamma function.

It is well known that, under the assumptions (A1)–(A2), $\{y_t, t \in \mathbb{Z}\}$ is a strictly stationary, invertible and square integrable process. Defining

$$\psi(z) = (1 - z)^{-d}\pi(z) = \sum_{j=0}^{\infty} \psi_j z^j, \qquad |z| < 1,$$

it is easily seen that $\{y_t, t \in \mathbb{Z}\}$ admits the causal moving-average representation

$$y_t - \mu = \psi(L)\varepsilon_t, \qquad t \in \mathbb{Z}, \tag{3}$$

with $\psi_0 = 1$. Thus, using Stirling's approximation formula, it may be shown that[2]

$$\psi_j \sim \{\pi(1)/\Gamma(d)\} j^{d-1} \qquad \text{as } j \to \infty,$$

from which it follows that $\sum_{j=0}^{\infty} \psi_j^2 < \infty$. Furthermore, letting $c(k) = \mathrm{Cov}(y_0, y_k)$ denote the autocovariance of $\{y_t, t \in \mathbb{Z}\}$ at lag $k \in \mathbb{Z}$, we have

$$c(k) \sim E(\varepsilon_0^2)\{\pi(1)/\Gamma(d)\}^2 B(d, 1 - 2d) k^{2d-1} \qquad \text{as } k \to \infty,$$

[2]Here and elsewhere, the symbol "$\sim$" indicates that the ratio of the left-hand and right-hand sides tends to 1 in the limit.

where $B(\cdot,\cdot)$ is the beta function. The series $\sum_{j=0}^{\infty}|\psi_j|$ and $\sum_{k=-\infty}^{\infty}|c(k)|$ are, therefore, properly divergent when $d \in (0, \frac{1}{2})$ and $\{y_t, t \in \mathbb{Z}\}$ is a process exhibiting strong dependence with memory (or fractional differencing) parameter d.

It is worth pointing out that the class of processes defined by (1)–(2) and (A1)–(A2) is rich enough to include many Gaussian and non-Gaussian strongly dependent processes. A prominent example are the popular fractional ARIMA processes introduced by Granger and Joyeux (1980) and Hosking (1981). For an ARIMA(p, d, q) process,

$$\pi(z) = \vartheta(z)/\varphi(z), \qquad |z| \leq 1, \tag{4}$$

with $\varphi(z) = 1 + \sum_{j=1}^{p}\varphi_j z^j$ and $\vartheta(z) = 1 + \sum_{j=1}^{q}\vartheta_j z^j$ being relatively prime finite-order polynomials having all their zeroes outside the closed disk $\{z \in \mathbb{C} : |z| \leq 1\}$. In this case, assumption (A2) is satisfied with $|\pi_j| = O(e^{-\beta j})$ as $j \to \infty$ for some $\beta \in (0, 1)$. Note, however, that the summability condition in (A2) also permits the weighting sequence $\{\pi_j, j \in \mathbb{Z}^+\}$ to decay at rates much slower than the exponential rate that is characteristic of ARMA processes (as is the case, for instance, when $\pi_j \sim K\, j^{-\kappa}$ as $j \to \infty$ for some $\kappa > 1$).[3]

3 Sieve Bootstrap: Definition and Properties

3.1 The Resampling Scheme

The sieve bootstrap scheme is motivated by the observation that, under assumption (A2), $\{y_t, t \in \mathbb{Z}\}$ admits the autoregressive representation

$$\phi(L)(y_t - \mu) = \varepsilon_t, \qquad t \in \mathbb{Z}, \tag{5}$$

where

$$\phi(z) = (1 - z)^d/\pi(z) = \sum_{j=0}^{\infty}\phi_j z^j, \qquad |z| < 1,$$

with $\phi_0 = 1$. Following Kreiss (1992) and Bühlmann (1997), the idea is to approximate (5) by a finite-order autoregressive model

$$\phi_h(L)(y_t - \mu) = \varepsilon_t^{(h)}, \qquad t \in \mathbb{Z}, \tag{6}$$

where, for some $h \in \mathbb{N}$,

$$\phi_h(z) = \sum_{j=0}^{h}\phi_j^{(h)} z^j, \qquad z \in \mathbb{C},$$

[3]Here and in the sequel, K denotes a generic finite constant whose value may change upon each appearance.

with $\phi_0^{(h)} = 1$. This autoregression can then be used to generated bootstrap replicates by means of a residual-based resampling plan. By allowing the order h of the autoregressive approximation to increase at some appropriate rate with the sample size, (6) may be interpreted as a sieve for the process defined by (5).

To give a formal definition of the sieve bootstrap, suppose that $\boldsymbol{y}_T = \{y_1, \ldots, y_T\}$ is a sample of size $T \in \mathbb{N}$ from $\{y_t, t \in \mathbb{Z}\}$ and let $h = h_T$ be a positive integer such that $h_T < T$ and $h_T \to \infty$ as $T \to \infty$. Further, let $\hat{\boldsymbol{\phi}}^{(h)} = (\hat{\phi}_1^{(h)}, \ldots, \hat{\phi}_h^{(h)})'$ be an estimator of the coefficient vector $\boldsymbol{\phi}^{(h)} = (\phi_1^{(h)}, \ldots, \phi_h^{(h)})'$ based on $\boldsymbol{y}_T$. It is well known that $\hat{\boldsymbol{\phi}}^{(h)}$ may be thought of as an estimator of the coefficients $\boldsymbol{\phi}^{(h)}$ of the best linear predictor of y_t based on $y_{t-1}, \ldots, y_{t-h}$, in which case the quantities $\boldsymbol{\phi}^{(h)}$ and $\sigma_h^2 = \mathrm{Var}(\varepsilon_t^{(h)})$ satisfy the Yule–Walker equations

$$\boldsymbol{C}^{(h)} \boldsymbol{\phi}^{(h)} = -\boldsymbol{c}^{(h)}, \tag{7}$$

$$\sigma_h^2 = c(0) + (\boldsymbol{\phi}^{(h)})' \boldsymbol{c}^{(h)}, \tag{8}$$

where $\boldsymbol{C}^{(h)} = \{c(i-j); i, j = 1, \ldots, h\}$ and $\boldsymbol{c}^{(h)} = (c(1), \ldots, c(h))'$ (cf. Brockwell and Davis (1991, p. 168)).

Defining the residuals associated with $\hat{\boldsymbol{\phi}}^{(h)}$ as

$$\hat{\varepsilon}_t = \sum_{j=0}^{h} \hat{\phi}_j^{(h)} (y_{t-j} - \bar{y}_T), \qquad h+1 \le t \le T,$$

with $\hat{\phi}_0^{(h)} = 1$, now put

$$\tilde{P}_T = (T-h)^{-1} \sum_{t=h+1}^{T} \delta_{\tilde{\varepsilon}_t},$$

where $\tilde{\varepsilon}_t = \hat{\varepsilon}_t - (T-h)^{-1} \sum_{t=h+1}^{T} \hat{\varepsilon}_t$ $(h+1 \le t \le T)$ and δ_x is the point mass at $x \in \mathbb{R}$. Then, the sieve bootstrap replicates $\{y_t^*, t \in \mathbb{Z}\}$ are defined via the recursive relation

$$\sum_{j=0}^{h} \hat{\phi}_j^{(h)} (y_{t-j}^* - \bar{y}_T) = \varepsilon_t^*, \qquad t \in \mathbb{Z}, \tag{9}$$

where $\{\varepsilon_t^*, t \in \mathbb{Z}\}$ is a sequence of conditionally i.i.d. random variables, given $\boldsymbol{y}_T$, with common distribution $\tilde{P}_T$. The sieve bootstrap version of any statistic $S_T = S_T(y_1, \ldots, y_T)$, which is a measurable function of $\boldsymbol{y}_T$, is given by $S_T^* = S_T(y_1^*, \ldots, y_T^*)$.

In practice, bootstrap replicates $(y_1^*, \ldots, y_T^*)$ may be obtained according to recursion (9) by setting $y_{-h+1}^* = \cdots = y_0^* = \bar{y}_T$, generating $T + q$ replicates with $q \in \mathbb{N}$ fairly large, and then discarding the initial q replicates to eliminate start-up effects. The order h of the autoregressive approximation may be selected adaptively by minimising (over a range of values of h) a model selection criterion such as the familiar AIC. For strongly dependent processes satisfying (1)–(2), Poskitt (2005) gives regularity conditions under which the autoregressive order selected through the AIC is asymptotically efficient in the sense of Shibata (1980).

3.2 Asymptotic Properties

In this subsection, we present some results on the structural properties of the sieve approximation in the form of lemmas. Some of the lemmas will be used subsequently but some are presented as they may be of independent interest in relation to the analysis of the autoregressive and moving-average representations of strongly dependent processes.

Letting the sample autocovariances of $\boldsymbol{y}_T$ be defined in the usual way as

$$c_T(k) = T^{-1} \sum_{t=1}^{T-|k|} (y_t - \bar{y}_T)(y_{t+|k|} - \bar{y}_T), \qquad k = 0, \pm 1, \ldots, \pm(T-1),$$

where $\bar{y}_T = T^{-1} \sum_{t=1}^{T} y_t$ is the sample mean, we make the following assumptions about the sieve bootstrap procedure.

(A3) $\{h = h_T, T \in \mathbb{N}\}$ is a sequence of positive integers such that $h_T \to \infty$ and $h_T = O\left(\{\ln T\}^{\alpha}\right)$ as $T \to \infty$ for some $\alpha \in (0, \infty)$.

(A4) $\hat{\boldsymbol{\phi}}^{(h)} = (\hat{\phi}_1^{(h)}, \ldots, \hat{\phi}_h^{(h)})'$ satisfies the empirical Yule–Walker equations

$$\boldsymbol{C}_T^{(h)} \hat{\boldsymbol{\phi}}^{(h)} = -\boldsymbol{c}_T^{(h)},$$

where $\boldsymbol{C}_T^{(h)} = \{c_T(i-j); i, j = 1, \ldots, h\}$ and $\boldsymbol{c}_T^{(h)} = (c_T(1), \ldots, c_T(h))'$.

Assumption (A3), which controls the rate of increase of the sieve order, is similar to assumptions that are often made in the theory of autoregressive approximations (see, e.g., An, Chen, and Hannan (1982)). It is worth noting that all our results concerning the asymptotic validity of the sieve bootstrap can in fact be obtained under the weaker condition $h_T = o(\{T/\ln T\}^{\frac{1}{2}-d})$ as $T \to \infty$, but we prefer to maintain (A3) because its requirement on the relative asymptotic rates of h and T does not depend on the unknown memory parameter d. Assumption (A4) guarantees that the polynomial

$$\hat{\phi}_h(z) = \sum_{j=0}^{h} \hat{\phi}_j^{(h)} z^j,$$

has no zeroes on the closed disk $\{z \in \mathbb{C} : |z| \leq 1\}$ (cf. Brockwell and Davis, 1991, p. 240). Hence, for each fixed $T \in \mathbb{N}$, the bootstrap process $\{y_t^*, t \in \mathbb{Z}\}$ defined by (9) admits the linear representation

$$y_t^* - \bar{y}_T = \hat{\psi}(L)\varepsilon_t^*, \qquad t \in \mathbb{Z}, \tag{10}$$

where

$$\hat{\psi}(z) = 1/\hat{\phi}_h(z) = \sum_{j=0}^{\infty} \hat{\psi}_j z^j, \qquad |z| \leq 1,$$

with $\hat{\psi}_0 = 1$.

The following lemma gives a uniform bound for the sequence $\{\hat{\psi}_j, j \in \mathbb{Z}^+\}$. (In the sequel, limits in order symbols are taken as $T \to \infty$, unless stated otherwise).

Lemma 1 *Let $\{y_t, t \in \mathbb{Z}\}$ satisfy* (1)–(2) *and suppose that assumptions* (A1)–(A4) *hold. Then*

$$\sup_{0 \leq j < \infty} \left| \hat{\psi}_j - \psi_j \right| = o_p(1). \tag{11}$$

Proof. Under the assumptions of the lemma, we have by Theorem 4.1 of Poskitt (2005) that

$$\max_{1 \leq j \leq h} |c_T(j) - c(j)| = O_p\left(\{\ln T/T\}^{\frac{1}{2}-d}\right) = o_p\left(\{\ln T\}^{\alpha}\right) \tag{12}$$

for every $\alpha > 0$. Coupling this with the rate of decay of $\{\phi_j, j \in \mathbb{Z}^+\}$, which is j^{-d-1} and hence faster than $(\ln j)^{\alpha}$, and by similar arguments to those used in the proof of Theorem 6 of An, Chen, and Hannan (1982), we deduce that

$$\max_{1 \leq j \leq h} \left| \hat{\phi}_j^{(h)} - \phi_j \right| = O_p\left(\{\ln T/T\}^{\frac{1}{2}-d}\right) = o_p\left(\{\ln T\}^{\alpha}\right). \tag{13}$$

Now, following Bühlmann (1995), let $\{x_t, t \in \mathbb{Z}\}$ be a process satisfying $\hat{\phi}_h(L)x_t = \eta_t$, where $\{\eta_t, t \in \mathbb{Z}\}$ is a sequence of centred i.i.d. random variables with $E(\eta_0^2) = E(\varepsilon_0^2) = \sigma^2$. It is easy to see that, for any integer $|k| \leq h$, we have $c^x(k) = \mathrm{Cov}(x_0, x_k) = c_T(k)\sigma^2/\hat{\sigma}^2$, where $\hat{\sigma}^2 = c_T(0) + (\hat{\boldsymbol{\phi}}^{(h)})'\boldsymbol{c}_T^{(h)}$.

Proceeding as in the proof of Theorem 3.2 of Bühlmann (1995), we first need to show that

$$\hat{\sigma}^2 = \sigma^2 + o_p(1). \tag{14}$$

We note that this result implies that $c^x(k) = O_p(1)$ for all $|k| \leq h$. From the Yule–Walker equations for σ^2 and $\hat{\sigma}^2$, we have

$$
\begin{aligned}
|\hat{\sigma}^2 - \sigma^2| &\le \left|\sum_{j=0}^{h}\left(\hat{\phi}_j^{(h)} c_T(j) - \phi_j c(j)\right)\right| + \left|\sum_{j=h+1}^{\infty} \phi_j c(j)\right| \\
&\le \max_{1\le j\le h}\left|\hat{\phi}_j^{(h)} - \phi_j\right| (h+1) \max_{1\le j\le h} |c_T(j) - c(j)| \sum_{j=1}^{h} |c(j)| \\
&+ \max_{1\le j\le h} |c_T(j) - c(j)| \sum_{j=0}^{h} |\phi_j| + |c(0)| \sum_{j=h+1}^{\infty} |\phi_j| \,. \qquad (15)
\end{aligned}
$$

By the assumptions of the lemma, (12) and (13), the first term on the r.h.s. of (15) is $o_p(1)$. A similar argument shows that the second term is also $o_p(1)$. For the third term, we note that $\phi_j \sim K j^{-d-1}$ as $j \to \infty$, which implies that

$$
\sum_{j=h+1}^{\infty} |\phi_j| = o(1) \quad \text{as } h \to \infty. \qquad (16)
$$

Thus, (14) holds.

Next, for any $j \in \mathbb{N}$,

$$
\begin{aligned}
\left|\hat{\psi}_j - \psi_j\right| &\le \sigma^{-2} \sum_{i=0}^{\infty} \left|\hat{\phi}_i^{(h)} - \phi_i\right| |c^x(j+i)| + \sigma^{-2} \sum_{i=0}^{\infty} |\phi_i| \, |c^x(j+i) - c(j+i)| \\
&= I_{j1} + I_{j2}. \qquad (17)
\end{aligned}
$$

Examining I_{j1} first, we have

$$
\sup_{j\in\mathbb{N}} I_{j1} \le \sigma^{-2} \max_{1\le i\le h} \left|\hat{\phi}_i^{(h)} - \phi_i\right| \sum_{i=0}^{h} |c^x(j+i)| + \sigma^{-2} \sum_{i=h+1}^{\infty} |\phi_i| \, |c^x(0)| \,. \qquad (18)
$$

But $\sum_{i=0}^{h} |c^x(j+i)| = O_p(\{\ln T\}^{\alpha})$ and $\max_{1\le i\le h} \left|\hat{\phi}_i^{(h)} - \phi_i\right| = o_p\left(\{\ln T\}^{\alpha}\right)$, so the first term on the r.h.s. of (18) is $o_p(1)$. Moreover, by (16) and the fact that $c^x(0) = O_p(1)$, the second term on the r.h.s. of (18) is also $o_p(1)$. For I_{j2}, we have

$$
\sup_{j\in\mathbb{N}} I_{j2} \le \sigma^{-2} \max_{1\le i\le h-j} |c_T(i) - c(i)| \sum_{i=0}^{h-j} |\phi_i| + \sigma^{-2} \max_{h-j+1\le i\le\infty} |c_T(i) - c(i)| \sum_{i=h-j+1}^{\infty} |\phi_i| \,. \qquad (19)
$$

But, since $\max_{1\le i\le h-j} |c_T(i) - c(i)| = o_p(1)$, $\max_{h-j+1\le i<\infty} |c_T(i) - c(i)| = O_p(1)$, $\sum_{i=0}^{h-j} |\phi_i| = O(1)$, and $\sum_{i=h-j+1}^{\infty} |\phi_i| = o(1)$ as $h \to \infty$, the r.h.s. of (19) is $o_p(1)$. Consequently, (11) holds. ■

Next, we establish some properties of the resampled innovation process $\{\varepsilon_t^*, t \in \mathbb{Z}\}$.[4]

Lemma 2 *Let* $\{y_t, t \in \mathbb{Z}\}$ *satisfy* (1)–(2) *and suppose that assumptions* (A1)–(A4) *hold. If* $E(|\varepsilon_0|^r) < \infty$ *for some* $r \geq \max\{2w, 4\}$*,* $w \in \mathbb{N}$*, then*

$$E^*(|\varepsilon_0^*|^{2w}) = E(|\varepsilon_0|^{2w}) + o_p(1). \tag{20}$$

Proof. To prove the lemma, we follow the steps of the proof of Lemma 5.3 of Bühlmann (1997). For ease of reference, we replicate the steps of that proof as most steps need to be modified somewhat. We start with the equality

$$E^*(|\varepsilon_0^*|^{2w}) = (T-h)^{-1} \sum_{t=h+1}^{T} \left(\hat{\varepsilon}_t - \hat{\varepsilon}_T^{(.)}\right)^{2w}, \tag{21}$$

where $\hat{\varepsilon}_T^{(.)} = (T-h)^{-1} \sum_{t=h+1}^{T} \hat{\varepsilon}_t$, and write

$$\hat{\varepsilon}_t = \varepsilon_t - (\bar{y}_T - \mu) \sum_{j=0}^{\infty} \phi_j + I_{t1} + I_{t2}, \tag{22}$$

with

$$I_{t1} = \sum_{j=0}^{h} \left(\hat{\phi}_j^{(h)} - \phi_j^{(h)}\right) (y_{t-j} - \bar{y}_T),$$

$$I_{t2} = \sum_{j=0}^{\infty} \left(\phi_j^{(h)} - \phi_j\right) (y_{t-j} - \bar{y}_T),$$

$(\phi_1^{(h)}, \ldots, \phi_h^{(h)})' = -[\boldsymbol{C}^{(h)}]^{-1} \boldsymbol{c}^{(h)}$, and $\phi_j^{(h)} = 0$ for $j > h$.

We first show that

$$\hat{\varepsilon}_T^{(.)} = o_p(1). \tag{23}$$

For this, observe that

$$\begin{aligned} \hat{\varepsilon}_T^{(.)} &= (T-h)^{-1} \left\{ \sum_{t=h+1}^{T} \left(\varepsilon_t - (\bar{y}_T - \mu) \sum_{j=0}^{\infty} \phi_j \right) + \sum_{t=h+1}^{T} I_{t1} + \sum_{t=h+1}^{T} I_{t2} \right\} \\ &= J_1 + J_2 + J_3. \end{aligned} \tag{24}$$

It is easy to see that $J_1 = o_p(1)$. By the Cauchy–Schwarz inequality, we have

$$|J_2| \leq \left(\sum_{j=0}^{h} \left(\hat{\phi}_j^{(h)} - \phi_j^{(h)} \right)^2 \right)^{\frac{1}{2}} \left((T-h)^{-1} \sum_{t=h+1}^{T} \sum_{j=0}^{h} (y_{t-j} - \bar{y}_T)^2 \right)^{\frac{1}{2}}.$$

[4] Henceforth, $\mathcal{L}^*(\cdot)$, $E^*(\cdot)$, $\mathrm{Var}^*(\cdot)$, and $\mathrm{Cov}^*(\cdot,\cdot)$ will be used to denote probability distribution, expectation, variance, and covariance, respectively, under the probability measure P^* induced by the resampling mechanism conditional on the original data $\boldsymbol{y}_T$.

But $(T-h)^{-1}\sum_{j=0}^{h}(y_{t-j}-\bar{y}_T)^2 = O_p(1)$ and so $(T-h)^{-1}\sum_{t=h+1}^{T}\sum_{j=0}^{h}(y_{t-j}-\bar{y}_T)^2 = O_p(h) = O_p(\{\ln T\}^{\alpha})$. Moreover, by virtue of Corollary 4.1 and Theorem 5.1 of Poskitt (2005), $\sum_{j=0}^{h}\left(\hat{\phi}_j^{(h)}-\phi_j^{(h)}\right)^2 = O_p\left(\{\ln T/T\}^{\frac{1}{2}-d}\right) = o_p\left(\{\ln T\}^{\alpha}\right)$ for all $\alpha > 0$. Hence, $J_2 = o_p(1)$. Finally, by Theorem 7.6.6 of Anderson (1971), I_{t2} converges to zero in mean square for all t and hence $J_3 = o_p(1)$. Note that this implies that

$$\sum_{j=0}^{\infty}\left(\phi_j^{(h)}-\phi_j\right) \leq K\sum_{j=h+1}^{\infty}\phi_j = o(1) \quad \text{as } h\to\infty. \tag{25}$$

Next, we show that

$$(T-h)^{-1}\sum_{t=h+1}^{T}\hat{\varepsilon}_t^{2w} = E\left(\varepsilon_0^{2w}\right) + o_p(1). \tag{26}$$

By using the Cauchy–Schwarz inequality and (25), we get

$$(T-h)^{-1}\sum_{t=h+1}^{T}|I_{t1}|^{2w} = O_p\left(\left\{(\ln T/T)^{\frac{1}{2}-d}\right\}^{2w}\right) = o_p(1) \tag{27}$$

and

$$(T-h)^{-1}\sum_{t=h+1}^{\infty}|I_{t2}|^{2w} = O_p\left(\left\{\sum_{j=h+1}^{\infty}|\phi_j|\right\}^{2w}\right) = o_p(1). \tag{28}$$

Hölder's inequality, (27) and (28) now yield (26). Finally, by a binomial expansion of (21), Hölder's inequality, (26) and (23), we obtain (21). ■

Lemma 3 *Let* $\{y_t, t\in\mathbb{Z}\}$ *satisfy* (1)–(2) *and suppose that assumptions* (A1)–(A4) *hold. Then, as* $T\to\infty$,

$$\mathcal{L}^*(\varepsilon_0^*) \to_w \mathcal{L}(\varepsilon_0) \quad \textit{in probability.}$$

Proof. The assertion of the lemma can be proved by arguing as in the proof of Lemma 5.4 of Bühlmann (1997) and using (23), (27), (28) and the fact that $\bar{y}_T-\mu = o_p(1)$. ■

4 Bootstrapping the Sample Mean

Let $\bar{y}_T^* = T^{-1}\sum_{t=1}^{T}y_t^*$ denote the bootstrap sample mean. Then, the conditional distribution of $\bar{y}_T^*-\bar{y}_T$ (suitably normalised), given $\boldsymbol{y}_T$, constitutes the sieve bootstrap approximation to the sampling distribution of $\bar{y}_T-\mu$ (suitably normalised).

Our first theorem establishes the asymptotic behaviour of the variance of the bootstrap sample mean. This result plays an important role in the proof of the consistency of the bootstrap estimator of the distribution of the sample mean but is also of interest in its own right.

Theorem 1 *Let* $\{y_t, t \in \mathbb{Z}\}$ *satisfy* (1)–(2) *and suppose that assumptions* (A1)–(A4) *hold. Then*

$$\mathrm{Var}^*(T^{\frac{1}{2}-d}\bar{y}_T^*) - \mathrm{Var}(T^{\frac{1}{2}-d}\bar{y}_T) = o_p(1).$$

Proof. Let $\{y_t^+, t \in \mathbb{Z}\}$ be the autoregressive process defined by the equation

$$\phi_h(L)(y_t^+ - \mu) = \sigma_h \eta_t^+, \qquad t \in \mathbb{Z},$$

where the parameters $\boldsymbol{\phi}^{(h)} = (\phi_1^{(h)}, \ldots, \phi_h^{(h)})'$ and σ_h^2 satisfy the Yule–Walker equations (7)–(8) and $\{\eta_t^+, t \in \mathbb{Z}\}$ is a sequence of i.i.d. random variables with $E(\eta_0^+) = 0$ and $E(|\eta_0^+|^2) = 1$. Writing $\bar{y}_T^+ = T^{-1}\sum_{t=1}^T y_t^+$, we must show that

$$\mathrm{Var}^*(T^{\frac{1}{2}-d}\bar{y}_T^*) - \mathrm{Var}(T^{\frac{1}{2}-d}\bar{y}_T^+) + \mathrm{Var}(T^{\frac{1}{2}-d}\bar{y}_T^+) - \mathrm{Var}(T^{\frac{1}{2}-d}\bar{y}_T) = o_p(1). \tag{29}$$

To this end, for any fixed $h, T \in \mathbb{N}$, put

$$\boldsymbol{\Phi}_T^{(h)} = \begin{bmatrix} -\phi_1^{(h)} & -\phi_2^{(h)} & -\phi_3^{(h)} & \cdots & -\phi_{T-1}^{(h)} & -\phi_T^{(h)} \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \tag{30}$$

with $\phi_j^{(h)} = 0$ for $j > h$, and define $\hat{\boldsymbol{\Phi}}_T^{(h)}$ to be the matrix obtained by replacing $\phi_j^{(h)}$ $(1 \le j \le h)$ by $\hat{\phi}_j^{(h)}$ in (30). Then, by taking into account Lemma 2 and the fact that $\sigma_h^2 \to E(\varepsilon_0^2)$ as $h \to \infty$, we have

$$\begin{aligned}
&\operatorname*{plim}_{T\to\infty} \sum_{k=1}^{T} \left\{\mathrm{Cov}^*(y_0^*, y_k^*) - \mathrm{Cov}(y_0^+, y_k^+)\right\} \\
&\le \operatorname*{plim}_{T\to\infty} \left\| E^*(|\varepsilon_0^*|^2)\left(\boldsymbol{I}_T - \hat{\boldsymbol{\Phi}}_T^{(h)} \otimes \hat{\boldsymbol{\Phi}}_T^{(h)}\right)^{-1} - \sigma_h^2\left(\boldsymbol{I}_T - \boldsymbol{\Phi}_T^{(h)} \otimes \boldsymbol{\Phi}_T^{(h)}\right)^{-1}\right\| \\
&\le \operatorname*{plim}_{T\to\infty} E(\varepsilon_0^2)\left\|\hat{\boldsymbol{\Phi}}_T^{(h)} \otimes \hat{\boldsymbol{\Phi}}_T^{(h)} - \boldsymbol{\Phi}_T^{(h)} \otimes \boldsymbol{\Phi}_T^{(h)}\right\|,
\end{aligned}$$

where $\boldsymbol{I}_T$ denotes the T-dimensional identity matrix, $\otimes$ is the Kronecker product operator, and $\|\boldsymbol{A}\| = \mathrm{tr}(\boldsymbol{A}'\boldsymbol{A})$ for a matrix $\boldsymbol{A}$. By using Corollary 4.1 and Theorem 5.1 of Poskitt (2005), we may, therefore, conclude that

$$\sum_{k=1}^{T} \left\{\mathrm{Cov}^*(y_0^*, y_k^*) - \mathrm{Cov}(y_0^+, y_k^+)\right\} = o_p(1). \tag{31}$$

In a similar fashion, defining $\mathbf{\Phi}_T$ to be the matrix obtained by replacing $\phi_j^{(h)}$ $(1 \le j \le T)$ by ϕ_j in (30), we have

$$
\begin{aligned}
&\lim_{T\to\infty} \sum_{k=1}^{T} \left\{ \mathrm{Cov}(y_0, y_k) - \mathrm{Cov}(y_0^+, y_k^+) \right\} \\
&\le \lim_{T\to\infty} \left\| E(\varepsilon_0^2) \left(\boldsymbol{I}_T - \mathbf{\Phi}_T \otimes \mathbf{\Phi}_T \right)^{-1} - \sigma_h^2 \left(\boldsymbol{I}_T - \mathbf{\Phi}_T^{(h)} \otimes \mathbf{\Phi}_T^{(h)} \right)^{-1} \right\| \\
&\le \lim_{T\to\infty} E(\varepsilon_0^2) \left\| \mathbf{\Phi}_T \otimes \mathbf{\Phi}_T - \mathbf{\Phi}_T^{(h)} \otimes \mathbf{\Phi}_T^{(h)} \right\| \\
&\le \lim_{T\to\infty} E(\varepsilon_0^2) \left\{ \left\| \mathbf{\Phi}_T \otimes \mathbf{\Phi}_T - \mathbf{\Phi}_T^{(h)} \otimes \mathbf{\Phi}_T \right\| + \left\| \mathbf{\Phi}_T^{(h)} \otimes \mathbf{\Phi}_T - \mathbf{\Phi}_T^{(h)} \otimes \mathbf{\Phi}_T^{(h)} \right\| \right\} \\
&\le 2 \lim_{T\to\infty} E(\varepsilon_0^2) \left\| \left(\mathbf{\Phi}_T - \mathbf{\Phi}_T^{(h)} \right) \otimes \mathbf{\Phi}_T \right\| \\
&\le \lim_{T\to\infty} E(\varepsilon_0^2) \left| \sum_{j=1}^{T} \left(\sum_{i=1}^{T} \left(\phi_i^{(h)} - \phi_i \right) \right) \phi_j \right|^2 \\
&\le \lim_{T\to\infty} E(\varepsilon_0^2) \left(\left| \sum_{j=1}^{h} \left(\phi_i^{(h)} - \phi_i \right) \right|^2 + \left| \sum_{j=h+1}^{T} \phi_j \right|^2 \right) \left| \sum_{j=1}^{T} \phi_j \right|^2 .
\end{aligned}
$$

Now, in view of Theorem 7.6.6 of Anderson (1971), $\sum_{j=1}^{h} \left(\phi_j^{(h)} - \phi_j \right) = o(1)$ as $h \to \infty$. Moreover, since $\lim_{T\to\infty} \sum_{j=1}^{T} \phi_j = O(1)$ and $\sum_{j=h+1}^{T} \phi_j = o(1)$, we conclude that

$$
\sum_{k=1}^{T} \left\{ \mathrm{Cov}(y_0, y_k) - \mathrm{Cov}(y_0^+, y_k^+) \right\} = o(1). \tag{32}
$$

Finally, it is easy to verify that

$$
\operatorname*{plim}_{T\to\infty} \left\{ \mathrm{Var}^*(T^{\frac{1}{2}-d} \bar{y}_T^*) - \mathrm{Var}(T^{\frac{1}{2}-d} \bar{y}_T^+) \right\} \le \operatorname*{plim}_{T\to\infty} T^{-2d} \sum_{k=1}^{T} \left\{ \mathrm{Cov}^*(y_0^*, y_k^*) - \mathrm{Cov}(y_0^+, y_k^+) \right\}
$$

and

$$
\lim_{T\to\infty} \left\{ \mathrm{Var}(T^{\frac{1}{2}-d} \bar{y}_T^+) - \mathrm{Var}(T^{\frac{1}{2}-d} \bar{y}_T) \right\} \le \lim_{T\to\infty} T^{-2d} \sum_{k=1}^{T} \left\{ \mathrm{Cov}(y_0^+, y_k^+) - \mathrm{Cov}(y_0, y_k) \right\},
$$

which, together with (31) and (32), prove (29) and thus the theorem. ∎

We are now ready to state the main result of this section, which shows that the sieve bootstrap approximation to the sampling distribution of $\bar{y}_T$ is asymptotically correct (to first order).

Theorem 2 *Let* $\{y_t, t \in \mathbb{Z}\}$ *satisfy* (1)–(2) *and suppose that assumptions* (A1)–(A4) *hold. Then*

$$\sup_{x\in\mathbb{R}} \left| P^*\left(T^{\frac{1}{2}-d}(\bar{y}_T^* - \bar{y}_T) \leq x\right) - P\left(T^{\frac{1}{2}-d}(\bar{y}_T - \mu) \leq x\right)\right| = o_p(1). \tag{33}$$

Proof. Arguing as in the proof of Theorem 2.1 of Wang, Lin, and Gulati (2003), it can be deduced that, under the assumptions of the theorem,

$$T^2 \mathrm{Var}\,(\bar{y}_T) \sim T^{1+2d}\omega^2 \qquad \text{as } T \to \infty, \tag{34}$$

where

$$\omega^2 = \frac{E(\varepsilon_0^2)\{\pi(1)\}^2 \varGamma(1-2d)}{(1+2d)\varGamma(1+d)\varGamma(1-d)}.$$

Then, an application of Theorem 18.6.5 of Ibragimov and Linnik (1971), in connection with (34) and Pólya's theorem, yields

$$\sup_{x\in\mathbb{R}} \left| P\left(T^{\frac{1}{2}-d}(\bar{y}_T - \mu) \leq x\right) - \varPhi(x/\omega)\right| = o(1),$$

where $\varPhi$ denotes the standard normal distribution function. Hence, in order to establish (33), it suffices to show that

$$\sup_{x\in\mathbb{R}} \left| P^*\left(T^{\frac{1}{2}-d}(\bar{y}_T^* - \bar{y}_T) \leq x\right) - \varPhi(x/\omega)\right| = o_p(1). \tag{35}$$

Now observe that

$$\mathrm{Var}^*(y_0^*) = c_T(0)E^*(|\varepsilon_0^*|^2)/\hat{\sigma}^2 = c_T(0)(1+o_p(1)) = O_p(1),$$

which, in view of (10) and Lemma 2, implies that

$$\sum_{j=0}^{\infty} \hat{\psi}_j^2 < \infty \quad \text{in probability} \tag{36}$$

for sufficiently large T. Therefore, recalling that, conditional on $\boldsymbol{y}_T$, $\{\varepsilon_t^*, t \in \mathbb{Z}\}$ is a collection of i.i.d. random variables with $E^*(\varepsilon_0^*) = 0$, it follows by Lemma 2, Theorem 1, (10), (34) and (36) that, for each subsequence $\{T_n\}$ of $\mathbb{N}$, there exists a further subsequence $\{T_m\} \subset \{T_n\}$ along which $\{(y_t^* - \bar{y}_T), t \in \mathbb{Z}\}$ satisfies almost surely all the conditions of Theorem 18.6.5 of Ibragimov and Linnik (1971). In consequence, by the bootstrap version of Slutsky's theorem and the continuity of $\varPhi$, we have

$$\lim_{m\to\infty} \sup_{x\in\mathbb{R}} \left| P^*\left(T_m^{\frac{1}{2}-d}(\bar{y}_{T_m}^* - \bar{y}_{T_m}) \leq x\right) - \varPhi(x/\omega)\right| = 0 \quad \text{a.s.},$$

from which (35) follows. ■

A difficulty that arises in the application of Theorem 1 for inference purposes in practice is that the rate of convergence $T^{\frac{1}{2}-d}$ depends on the unknown memory parameter d. However, if an estimator $\hat{d} = \hat{d}(y_1, \ldots, y_T)$ of the memory parameter is available which converges to d sufficiently fast as $T \to \infty$, then the problem may be overcome by using the estimated rate of convergence $T^{\frac{1}{2}-\hat{d}}$ in lieu of $T^{\frac{1}{2}-d}$. The following extension of Theorem 2 shows that the asymptotic validity of the sieve bootstrap is not affected by substituting a suitable estimator of d in the scaling factor $T^{\frac{1}{2}-d}$.

Corollary 1 *Let* $\{y_t, t \in \mathbb{Z}\}$ *satisfy* (1)–(2) *and suppose that assumptions* (A1)–(A4) *hold. If* $\hat{d} = d + o_p(1/\ln T)$*, then*

$$\sup_{x \in \mathbb{R}} \left| P^* \left(T^{\frac{1}{2}-\hat{d}} (\bar{y}_T^* - \bar{y}_T) \leq x \right) - P \left(T^{\frac{1}{2}-d} (\bar{y}_T - \mu) \leq x \right) \right| = o_p(1). \tag{37}$$

Proof. The claim (37) follows from Theorem 2 and the fact that $T^{d-\hat{d}} = 1 + o_p(1)$ as a result of $\hat{d} - d = o_p(1/\ln T)$. ■

It is worth mentioning that estimators of d satisfying the requirement of Corollary 1 are readily available, such as semiparametric log-periodogram and local Whittle estimators.

5 Bootstrapping the Sample Autocovariances

In this section, we examine the properties of the sieve bootstrap estimator of the distribution of sample autocovariances when the latter are asymptotically normal with an $O_p(1/\sqrt{T})$ convergence rate. As is well known, for a strongly dependent process satisfying (1)–(2) and (A1)–(A2), this is the case when $d < \frac{1}{4}$ (cf. Hosking (1996)).

For each $T \in \mathbb{N}$, the bootstrap sample autocovariances are defined as

$$c_T^*(k) = T^{-1} \sum_{t=1}^{T-|k|} (y_t^* - \bar{y}_T^*)(y_{t+|k|}^* - \bar{y}_T^*), \qquad k = 0, \pm 1, \ldots, \pm(T-1).$$

Then, the conditional distribution of $\sqrt{T}\{c_T^*(k) - c^*(k)\}$, given $\boldsymbol{y}_T$, provides the bootstrap approximation to the sampling distribution of $\sqrt{T}\{c_T(k) - c(k)\}$, where $c^*(k) = \mathrm{Cov}^*(y_0^*, y_k^*)$. The following theorem shows that such an approximation is asymptotically correct.

Theorem 3 *Let $\{y_t, t \in \mathbb{Z}\}$ satisfy* (1)–(2) *with $d \in (0, \frac{1}{4})$ and suppose that assumptions* (A1)–(A4) *hold. Then, for any $k \in \mathbb{Z}$ with $|k| < T$,*

$$\sup_{x \in \mathbb{R}} \left| P^* \left(\sqrt{T}\{c_T^*(k) - c^*(k)\} \leq x \right) - P\left(\sqrt{T}\{c_T(k) - c(k)\} \leq x \right) \right| = o_p(1). \tag{38}$$

Proof. By reasoning along the lines of the proof of Theorem 2, assertion (38) follows from the central limit theorem of Hannan (1976) after verifying that

$$\sum_{k=1}^{\infty} \left\{ [\mathrm{Cov}^*(y_0^*, y_k^*)]^2 - [\mathrm{Cov}(y_0^+, y_k^+)]^2 \right\} = o_p(1) \tag{39}$$

and

$$\sum_{k=1}^{\infty} \left\{ [\mathrm{Cov}(y_0, y_k)]^2 - [\mathrm{Cov}(y_0^+, y_k^+)]^2 \right\} = o(1), \tag{40}$$

where $\{y_t^+, t \in \mathbb{Z}\}$ is the autoregressive process defined in the proof of Theorem 1. Focusing on (39) first, and using the same notation as in the proof of Theorem 1, we have

$$\begin{aligned}
&\operatorname*{plim}_{T\to\infty} \sum_{k=1}^{T} \left(\{\mathrm{Cov}^*(y_0^*, y_k^*)\}^2 - \{\mathrm{Cov}(y_0^+, y_k^+)\}^2 \right) \\
&\leq \operatorname*{plim}_{T\to\infty} \left\| \left[E^*(|\varepsilon_0^*|^2) \left(\boldsymbol{I}_T - \hat{\boldsymbol{\Phi}}_T^{(h)} \otimes \hat{\boldsymbol{\Phi}}_T^{(h)} \right)^{-1} \right]^{(2)} - \left[\sigma_h^2 \left(\boldsymbol{I}_T - \boldsymbol{\Phi}_T^{(h)} \otimes \boldsymbol{\Phi}_T^{(h)} \right)^{-1} \right]^{(2)} \right\| \\
&\leq \operatorname*{plim}_{T\to\infty} \left\{ \left\| E^*(|\varepsilon_0^*|^2) \left(\boldsymbol{I}_T - \hat{\boldsymbol{\Phi}}_T^{(h)} \otimes \hat{\boldsymbol{\Phi}}_T^{(h)} \right)^{-1} - \sigma_h^2 \left(\boldsymbol{I}_T - \boldsymbol{\Phi}_T^{(h)} \otimes \boldsymbol{\Phi}_T^{(h)} \right)^{-1} \right\| \right. \\
&\qquad \left. \times \left\| \sigma_h^2 \left(\boldsymbol{I}_T - \boldsymbol{\Phi}_T^{(h)} \otimes \boldsymbol{\Phi}_T^{(h)} \right)^{-1} \right\| \right\},
\end{aligned} \tag{41}$$

where $\boldsymbol{A}^{(r)}$ denotes the rth Hadamard power of the matrix $\boldsymbol{A}$. Given that $\sigma_h^2 \to E(\varepsilon_0^2)$ as $h \to \infty$, relation (41), together with Lemma 2, Corollary 4.1 and Theorem 5.1 of Poskitt (2005), implies (39). A similar argument may be used to show that (40) holds. ■

It is straightforward to infer that, under the conditions of Theorem 3, the bootstrap approximation to the distribution of $\sqrt{T}(c_T(0) - c(0), c_T(1) - c(1), \ldots, c_T(n) - c(n))'$ is consistent for any fixed $n \in \mathbb{Z}^+$. This result may in turn be used to establish consistency of the bootstrap approximation to the distribution of functions of sample autocovariances. An example is the minimum-distance estimator of the memory parameter d proposed by Tieslau, Schmidt, and Baillie (1996), the small-sample properties of which will be examined in the next section of the paper.

6 Numerical Evidence

This section discusses some simulation experiments that illustrate the finite-sample performance of the sieve bootstrap for strongly dependent process. We present two sets of experiments, one for the case of the sample mean and one for the minimum-distance estimator of the memory parameter.

6.1 Sample Mean

The data-generating mechanism used in the experiments is the ARIMA$(1, d, 1)$ model (4) with $(\varphi_1, \vartheta_1) = (0, 0)$, $(\varphi_1, \vartheta_1) = (-0.3, 0.4)$, or $(\varphi_1, \vartheta_1) = (0.3, -0.4)$. In each case, the distribution of ε_0 is standard normal, $d \in \{0.1, 0.2\}$, and $\mu = 0$.

In Table 1, we report the mean, variance, skewness and kurtosis of the distribution of the sample mean, along with their bootstrap estimates. The former are computed from 1000 Monte Carlo values of $T^{\frac{1}{2}-d}(\bar{y}_T - \mu)$. The bootstrap estimates are computed as averages over 1000 Monte Carlo trials of the moments of 199 replicates of $T^{\frac{1}{2}-\hat{d}}(\bar{y}^*_T - \bar{y}_T)$. The memory parameter d is estimated by using the bias-reduced log-periodogram regression estimator of Andrews and Guggenberger (2003) (with $r = 2$ and $m = \lfloor T^{0.8} \rfloor$, in their notation).[5] The sample size and autoregressive sieve order, both in this and the next subsection, are chosen to be $T \in \{100, 200, 400\}$ and $h = \lfloor (\ln T)^2 \rfloor$.

The simulation results show that the sieve bootstrap approximation to the first four moments of the finite-sample distribution of the normalised sample mean is quite accurate, particularly when $T = 400$. The bootstrap approximation tends to have larger variance than the distribution of the sample mean for the smaller sample sizes.

6.2 Minimum-Distance Estimator

The minimum-distance estimator of Tieslau, Schmidt, and Baillie (1996) minimises the distance between the sample and theoretical autocorrelations of a strongly dependent process. In the case of an ARIMA$(0, d, 0)$ process, on which we focus, the minimum-distance estimator $\hat{d}$ of d is defined as

$$\hat{d} = \arg\min_d \left(\boldsymbol{\rho}_T^{(n)} - \boldsymbol{\rho}^{(n)}\right)' \boldsymbol{W} \left(\boldsymbol{\rho}_T^{(n)} - \boldsymbol{\rho}^{(n)}\right),$$

where, for some fixed $n \in \mathbb{N}$, $\boldsymbol{\rho}^{(n)} = (\rho(1), \dots, \rho(n))'$, $\boldsymbol{\rho}_T^{(n)} = (\rho_T(1), \dots, \rho_T(n))'$,

$$\rho(j) = \frac{\Gamma(1-d)\Gamma(j+d)}{\Gamma(d)\Gamma(j-d+1)}, \qquad j \in \mathbb{N},$$

[5] Here, we write $\lfloor x \rfloor$ to denote the integer part of the real number x.

$$\rho_T(j) = c_T(j)/c_T(0), \qquad |j| < T,$$

and $\boldsymbol{W}$ is a $n \times n$ symmetric, positive definite weighting matrix. When $d \in (-\frac{1}{2}, \frac{1}{4})$, the asymptotic distribution of $\sqrt{T}(\hat{d} - d)$ is normal with mean zero and covariance matrix $(\boldsymbol{b}'\boldsymbol{W}\boldsymbol{b})^{-1}\boldsymbol{b}'\boldsymbol{W}\boldsymbol{V}\boldsymbol{W}\boldsymbol{b}(\boldsymbol{b}'\boldsymbol{W}\boldsymbol{b})^{-1}$, where $\boldsymbol{b} = \partial\boldsymbol{\rho}^{(n)}/\partial d$ and $\boldsymbol{V} = [v_{ij}]$ is the asymptotic covariance matrix of $\sqrt{T}\left(\boldsymbol{\rho}_T^{(n)} - \boldsymbol{\rho}^{(n)}\right)$,

$$v_{ij} = \sum_{s=1}^{\infty}\{\rho(s+i) + \rho(s-i) - 2\rho(i)\rho(s)\}\{\rho(s+j) + \rho(s-j) - 2\rho(j)\rho(s)\}, \quad i, j = 1, \ldots, n.$$

Table 2 reports the exact moments of the minimum-distance estimator, computed from 1000 Monte Carlo replications, and the average bootstrap estimates of these moments based on 199 bootstrap replications. The data-generating mechanism is a Gaussian ARIMA$(0, d, 0)$ process, and $\hat{d}$ is computed by setting $n = \lfloor T/10 \rfloor$ and $\boldsymbol{W} = \boldsymbol{I}_n$ (this choice for the weighting matrix is made to ease computation). The finite-sample moments of the estimator are well approximated by their sieve bootstrap estimates even for the smaller sample sizes.

7 Some Further Observations

We end with two final observations on the asymptotic results obtained in Sections 3–5.

First, the assumption that $\{\varepsilon_t, t \in \mathbb{Z}\}$ is a sequence of i.i.d. random variables can be relaxed in certain circumstances, namely in Lemmas 1–3 and Theorem 3, without any essential changes in the proofs. To be more specific, (A1) may be replaced with the assumption that $\{\varepsilon_t, t \in \mathbb{Z}\}$ is a strictly stationary and ergodic sequence such that $E(\varepsilon_t|\mathcal{F}_{t-1}) = 0$ a.s., $E(\varepsilon_t^2|\mathcal{F}_{t-1}) = \sigma^2$ a.s., and $E(\varepsilon_t^4) < \infty$, where $\mathcal{F}_t$ is the σ-algebra generated by $\{\varepsilon_s, s \leq t\}$.

Second, although we have assumed throughout that $d > 0$, because of our interest in strongly dependent processes, the conclusions of Theorems 1–3 remain valid when $\{y_t, t \in \mathbb{Z}\}$ satisfies (1)–(2) with $d \in (-\frac{1}{2}, 0]$. If $d = 0$, then $\{y_t, t \in \mathbb{Z}\}$ is evidently a weakly dependent linear process such that $\sum_{j=0}^{\infty}|\psi_j| < \infty$ and $\sum_{k=-\infty}^{\infty}c(k) > 0$. When $d \in (-\frac{1}{2}, 0)$, $\{y_t, t \in \mathbb{Z}\}$ is usually said to be anti-persistent; in this case, $\sum_{j=0}^{\infty}\psi_j = \sum_{k=-\infty}^{\infty}c(k) = 0$, but, unlike the case $d \in (0, \frac{1}{2})$, the series $\sum_{k=-\infty}^{\infty}c(k)$ is absolutely convergent, even though $c(k)$ tends to zero at a hyperbolic rate as $k \to \infty$.

References

An, H.-Z., Z.-G. Chen, and E. J. Hannan (1982): "Autocorrelation, Autoregression and Autoregressive Approximation," *Annals of Statistics*, 10, 926–936 (Correction, 11, 1018).

Anderson, T. W. (1971): *The Statistical Analysis of Time Series*. Wiley, New York.

Andrews, D. W. K., and P. Guggenberger (2003): "A Bias-Reduced Log-Periodogram Regression Estimator for the Long-Memory Parameter," *Econometrica*, 71, 675–712.

Andrews, D. W. K., and O. Lieberman (2002): "Higher-Order Improvements of the Parametric Bootstrap for Long-Memory Gaussian Processes," Cowles Foundation Discussion Paper No. 1378, Yale University.

Beran, J. (1994): *Statistics for Long-Memory Processes*. Chapman and Hall, London.

Bickel, P. J., and P. Bühlmann (1996): "What is a Linear Process?," *Proceedings of the National Academy of Sciences of the USA*, 93, 12128–12131.

——— (1997): "Closure of Linear Processes," *Journal of Theoretical Probability*, 10, 445–479.

——— (1999): "A New Mixing Notion and Functional Central Limit Theorems for a Sieve Bootstrap in Time Series," *Bernoulli*, 5, 413–446.

Brockwell, P. J., and R. A. Davis (1991): *Time Series: Theory and Methods*. 2nd Edition, Springer–Verlag, New York.

Bühlmann, P. (1995): "Moving-Average Representation of Autoregressive Approximations," *Stochastic Processes and Their Applications*, 60, 331–342.

——— (1997): "Sieve Bootstrap for Time Series," *Bernoulli*, 3, 123–148.

Choi, E., and P. Hall (2000): "Bootstrap Confidence Regions Constructed from Autoregressions of Arbitrary Order," *Journal of the Royal Statstical Society B*, 62, 461–477.

Granger, C. W. J., and R. Joyeux (1980): "An Introduction to Long-Memory Time Series Models and Fractional Differencing," *Journal of Time Series Analysis*, 1, 15–29.

Hall, P., B.-Y. Jing, and S. N. Lahiri (1998): "On the Sampling Window Method for Long-Range Dependent Data," *Statistica Sinica*, 8, 1189–1204.

Hannan, E. J. (1976): “The Asymptotic Distribution of Serial Covariances,” *Annals of Statistics*, 4, 396–399.

Hidalgo, J. (2003): “An Alternative Bootstrap to Moving Blocks for Time Series Regression Models,” *Journal of Econometrics*, 117, 369–399.

Hosking, J. R. M. (1981): “Fractional Differencing,” *Biometrika*, 65, 165–176.

——— (1996): “Asymptotic Distributions of the Sample Mean, Autocovariances, and Autocorrelations of Long-Memory Time Series,” *Journal of Econometrics*, 73, 261–284.

Ibragimov, I. A., and Y. V. Linnik (1971): *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff, Groningen.

Kreiss, J.-P. (1992): “Bootstrap Procedures for AR(∞) Processes,” in *Bootstrapping and Related Techniques*, ed. by K.-H. Jöckel, G. Rothe, and W. Sendler, no. 376 in Lecture Notes in Economics and Mathematical Systems, pp. 107–113. Springer–Verlag, Heidelberg.

Lahiri, S. N. (1993): “On the Moving Block Bootstrap under Long Range Dependence,” *Statistics and Probability Letters*, 18, 405–413.

Nordman, D. J., and S. N. Lahiri (2005): “Validity of the Sampling Window Method for Long-Range Dependent Linear Processes,” *Econometric Theory*, 21, 1087–1111.

Paparoditis, E. (1996): “Bootstrapping Autoregressive and Moving Average Parameter Estimates of Infinite Order Vector Autoregressive Processes,” *Journal of Multivariate Analysis*, 57, 277–296.

Poskitt, D. S. (2005): “Autoregressive Approximation in Nonstandard Situations: The Non-Invertible and Fractionally Integrated Cases,” Working Paper 16/05, Department of Econometrics and Business Statistics, Monash University.

Shibata, R. (1980): “Asymptotically Efficient Selection of the Order of the Model for Estimating Parameters of a Linear Process,” *Annals of Statistics*, 8, 147–164.

Tieslau, M. A., P. Schmidt, and R. T. Baillie (1996): “A Minimum Distance Estimator for Long-Memory Processes,” *Journal of Econometrics*, 71, 249–264.

Wang, Q., Y.-X. Lin, and C. M. Gulati (2003): “Asymptotics for General Fractionally Integrated Processes with Applications to Unit Root Tests,” *Econometric Theory*, 19, 143–164.

Table 1. Distribution of the Normalised Sample Mean									
d	T	Exact Moments				Bootstrap Estimates			
		Mean	Variance	Skew	Kurtosis	Mean	Variance	Skew.	Kurtosis
		$\varphi_1 = 0, \quad \vartheta_1 = 0$							
0.1	100	-0.04642	0.91175	-0.07685	3.20484	0.00147	2.10066	0.00278	2.93953
	200	-0.02148	0.95351	-0.05424	3.08918	0.00296	2.01207	-0.00219	2.94836
	400	-0.02066	0.96933	-0.10287	2.81121	-0.00077	1.64797	-0.00095	2.94906
0.2	100	0.06227	0.90547	-0.11519	3.00027	0.00107	2.02144	-0.00273	2.93361
	200	0.01462	0.96454	-0.07740	2.96915	0.00368	1.46138	-0.00820	2.94306
	400	0.01928	0.88641	-0.03137	2.88946	-0.00143	0.87627	-0.00152	2.95700
		$\varphi_1 = -0.3, \quad \vartheta_1 = 0.4$							
0.1	100	-0.06526	3.88057	-0.04235	2.92432	0.00160	6.46742	-0.00219	2.95105
	200	-0.00720	4.00673	-0.05891	2.84396	0.00239	5.55128	0.00561	2.95731
	400	-0.05300	3.84339	0.03108	2.80514	0.00233	4.50300	0.00068	2.95261
0.2	100	0.07912	3.82757	-0.05133	2.94615	-0.00100	6.20345	-0.00478	2.93877
	200	0.06970	3.73352	0.01083	2.90068	0.00346	4.63784	-0.00423	2.95363
	400	0.01138	3.59040	-0.04309	3.04852	-0.00907	3.14256	-0.00091	2.94150
		$\varphi_1 = 0.3, \quad \vartheta_1 = -0.4$							
0.1	100	-0.00501	0.20556	-0.04549	2.92632	0.00047	0.52266	-0.00407	2.95074
	200	0.00440	0.21568	0.035061	3.02836	0.00065	0.49768	0.00377	2.93545
	400	-0.01831	0.21165	-0.01803	2.98310	0.00235	0.30666	0.00016	2.92061
0.2	100	-0.00980	0.21491	0.00997	2.95249	-0.00083	0.27508	0.00014	2.94270
	200	0.00097	0.20401	0.12586	2.92149	-0.00049	0.28233	0.00227	2.93752
	400	-0.01442	0.20909	-0.00872	2.92541	0.00144	0.19989	0.00043	2.93173

Table 2. Distribution of the Minimum-Distance Estimator of d									
d	T	Exact Moments				Bootstrap Estimates			
		Mean	Variance	Skew.	Kurtosis	Mean	Variance	Skew.	Kurtosis
0.1	100	0.08074	0.00214	-0.15803	3.07319	0.07748	0.00251	-0.13405	3.12112
	200	0.08533	0.00191	0.16460	3.06314	0.08201	0.00202	-0.01210	3.04369
	400	0.08350	0.00118	0.13181	3.21260	0.08247	0.00136	0.07634	3.06973
0.2	100	0.16534	0.00230	-0.15513	3.01301	0.16436	0.00251	-0.09461	2.88084
	200	0.16973	0.00224	0.12000	2.57694	0.16745	0.00221	0.00857	2.83541
	400	0.16910	0.00162	0.18399	2.83958	0.16894	0.00172	0.10313	2.81464

Queen Mary
University of London

This working paper has been produced by
the Department of Economics at
Queen Mary, University of London

Department of Economics
Queen Mary, University of London
Mile End Road
London E1 4NS
Tel: +44 (0)20 7882 5096
Fax: +44 (0)20 8983 3580
Web: www.econ.qmul.ac.uk/papers/wp.htm