



MONASH University

Australia

Department of Econometrics and Business Statistics

<http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/>

**Higher-Order Improvements of the Sieve Bootstrap
For Fractionally Integrated Processes**

D.S. Poskitt, Simone D. Grose and Gael M. Martin

November 2013

Working Paper 25/13 (Revised 09/12)

Higher-Order Improvements of the Sieve Bootstrap for Fractionally Integrated Processes*

D. S. Poskitt, Simone D. Grose and Gael M. Martin[†]

Department of Econometrics & Business Statistics, Monash University

October 15, 2013

Abstract

This paper investigates the accuracy of bootstrap-based inference in the case of long memory fractionally integrated processes. The re-sampling method is based on the semi-parametric sieve approach, whereby the dynamics in the process used to produce the bootstrap draws are captured by an autoregressive approximation. Application of the sieve method to data pre-filtered by a semi-parametric estimate of the long memory parameter is also explored. Higher-order improvements yielded by both forms of re-sampling are demonstrated using Edgeworth expansions for a broad class of statistics that includes first- and second-order moments, the discrete Fourier transform and regression coefficients. The methods are then applied to the problem of estimating the sampling distributions of the sample mean and of selected sample autocorrelation coefficients, in experimental settings. In the case of the sample mean, the pre-filtered version of the bootstrap is shown to avoid the distinct underestimation of the sampling variance of the mean which the raw sieve method demonstrates in finite samples, higher order accuracy of the latter notwithstanding. Pre-filtering also produces gains in terms of the accuracy with which the sampling distributions of the sample autocorrelations are reproduced, most notably in the part of the parameter space in which asymptotic normality does not obtain. Most importantly, the sieve bootstrap is shown to reproduce the (empirically infeasible) Edgeworth expansion of the sampling distribution of the autocorrelation coefficients, in the part of the parameter space in which the expansion is valid.

*This research has been supported by Australian Research Council (ARC) Discovery Grants DP0985234 and DP120102344, and ARC Future Fellowship FT0991045. The authors would like to thank the Editor and two referees for very detailed and constructive comments on an earlier draft of the paper.

[†]Corresponding author: Gael Martin, Department of Econometrics and Business Statistics, Monash University, Clayton, Victoria 3800, Australia. Tel.: +61-3-9905-1189; fax: +61-3-9905-5474; email: gael.martin@monash.edu.

Keywords: Long memory, ARFIMA, sieve bootstrap, bootstrap-based inference, Edgeworth expansion, sampling distribution.

JEL Classification: C18, C22, C52

1 Introduction

Many empirical time series have been found to exhibit behaviour characteristic of long memory, or long-range dependent, processes, and the class of fractionally integrated ($I(d)$) processes introduced by [Granger and Joyeux \(1980\)](#) and [Hosking \(1980\)](#) is perhaps the most popular model used to describe the features of such processes. $I(d)$ processes can be characterized by the specification

$$y(t) = \sum_{j=0}^{\infty} k(j)\varepsilon(t-j) = \frac{\kappa(z)}{(1-z)^d} \varepsilon(t), \quad (1.1)$$

where $\varepsilon(t)$, $t \in \mathbb{Z}$, is a zero mean white noise process with variance σ^2 , z is here interpreted as the lag operator ($z^j y(t) = y(t-j)$), and $\kappa(z) = \sum_{j \geq 0} \kappa(j)z^j$. The behaviour of this process naturally depends on the fractional integration parameter d ; for instance, if the “non-fractional” component $\kappa(z)$ is the transfer function of a stable, invertible autoregressive moving-average (ARMA) and $|d| < 0.5$, then the coefficients of $k(z)$ are square-summable, $\sum_{j \geq 0} |k(j)|^2 < \infty$, and $y(t)$ is well-defined as the limit in mean square of a covariance-stationary process. More pertinently, for any $d > 0$ the impulse response coefficients of $k(z)$ in the representation (1.1) are not absolutely summable and the autocovariances decline at a hyperbolic rate, $\gamma(\tau) \sim C\tau^{2d-1}$, rather than the exponential rate typical of an ARMA process. For a detailed description of the properties of long memory processes see [Beran \(1994\)](#).

Statistical procedures for analyzing fractional processes are discussed in [Hosking \(1996\)](#), and techniques for estimating fractional models have ranged from the likelihood-based methods studied in [Fox and Taquq \(1986\)](#), [Dahlhaus \(1989\)](#), [Sowell \(1992\)](#) and [Beran \(1995\)](#), to the semi-parametric methods advanced by [Geweke and Porter-Hudak \(1983\)](#) and [Robinson \(1995a,b\)](#), among others. These techniques typically focus on obtaining an accurate estimate of the parameter governing the long-term behaviour of the process, and the asymptotic theory for these estimators is well established. In particular, we have consistency, asymptotic efficiency, and asymptotic normality for the maximum likelihood estimator (MLE), and the semi-parametric estimators are consistent and asymptotically pivotal with particularly simple asymptotic normal distributions.

Concurrent with the development of the asymptotic theory associated with the estimation of long memory models, focus has also been directed at the production of more accurate estimates of finite sample distributions in this setting. An explicit form for the Edgeworth expansion for the sample autocorrelation function of a stationary Gaussian long memory process is derived in [Lieberman, Rousseau and Zucker \(2001\)](#), and [Lieberman, Rousseau and Zucker \(2003\)](#) establish the validity of an Edgeworth expansion for the distribution of the MLE of the parameters of such a process, with a

zero mean assumed. The unknown mean case is covered in [Andrews and Lieberman \(2005\)](#), with the estimator defined by maximizing the log-likelihood with the unknown mean replaced by the sample mean (referred to as the “plug-in” MLE, or PML). [Andrews and Lieberman \(2005\)](#) also derive results for the Whittle MLE (WML) and for the plug-in version (PWML). [Giraitis and Robinson \(2003\)](#) derive an Edgeworth expansion for the semi-parametric local Whittle estimator of the long memory parameter ([Robinson, 1995a](#)) (SPLW), whilst [Lieberman and Phillips \(2004\)](#) derive an explicit form for the first-order expansion for the MLE of the long memory parameter in the fractional noise case.

From the point of view of practical implementation, evaluation of the terms in such expansions, for general long memory models, is no trivial task and typically requires knowledge of the values of population ensemble parameters. These expansions are also usually only valid under more restrictive assumptions than are required for first-order asymptotic approximations; see, for example, [Lieberman et al. \(2001\)](#) and [Giraitis and Robinson \(2003\)](#). Accordingly, much attention has also been given to the application of bootstrap-based inference in these models. Building on the Edgeworth results of [Lieberman et al. \(2003\)](#) and [Andrews and Lieberman \(2005\)](#), [Andrews, Lieberman and Marmer \(2006\)](#) derive the error rate for the parametric bootstrap for the PML and PWML estimators in Gaussian autoregressive fractionally integrated moving average (ARFIMA) models. In contrast, [Poskitt \(2008\)](#) proposes a semi-parametric approach, based on the sieve bootstrap, and provides both theoretical and simulation-based results regarding the accuracy with which the method estimates the true sampling distribution of suitably continuous linear statistics. To the authors’ knowledge [Andrews et al. \(2006\)](#) and [Poskitt \(2008\)](#) are amongst the earliest papers in the literature to have examined the theoretical properties of bootstrap methods in the context of fractionally integrated (long memory) processes.

The current paper builds upon the results presented in [Poskitt \(2008\)](#) and produces new results regarding error rates for sieve-based bootstrap techniques in the context of fractionally integrated processes. Using Edgeworth expansions, it is shown that the procedure we here refer to as the “raw” sieve bootstrap can achieve an error rate of $O_p(T^{-(1-d')+\beta})$ for all $\beta > 0$ where $d' = \max\{0, d\}$, for a class of statistics that includes the sample mean, the sample autocovariance and autocorrelation functions, the discrete Fourier transform and ordinary least squares (OLS) regression coefficients. We also present a new methodology based on a modified form of the sieve bootstrap. The modification uses a consistent semi-parametric estimator of the long memory parameter to pre-filter the raw data, prior to the application of a long autoregressive approximation which acts as the “sieve” from which bootstrap samples are produced. We refer to this as the pre-filtered sieve bootstrap. We establish that, subject to appropriate regularity, for any fractionally integrated processes with $|d| < 0.5$ the er-

ror rate of the pre-filtered sieve bootstrap is $O_p(T^{-1+\beta})$ for all $\beta > 0$. These results generalize those of [Choi and Hall \(2000\)](#) who show that, for linear statistics characterized by polynomial products, double sieve bootstrap calibrated percentile methods and sieve bootstrap percentile t confidence intervals evaluated in the short memory case converge at a rate arbitrarily close to that obtained with simple random samples, namely $O_p(T^{-1+\beta})$ for all $\beta > 0$.

[Choi and Hall \(2000\)](#) argue that for short memory processes the sieve bootstrap is to be preferred over the block bootstrap ([Künsch, 1989](#)). In particular they note that although the block bootstrap accurately replicates the first-order dependence structure of the original times series it fails to reproduce second-order effects, because these are corrupted by the blocking process. Use of an adjusted variance estimate to correct for the failure to approximate second-order effects results, in turn, in an error rate of only $O_p(T^{-2/3+\beta})$ for the block bootstrap. In contrast, the second-order structure is shown to be preserved by the sieve. [Choi and Hall \(2000\)](#) demonstrate that the performance of the sieve is robust to the selected order for the autoregressive approximation, whilst noting that the choice of block length and other tuning parameters can be crucial to the performance of the block bootstrap. Moreover, as these authors also remark, the use of an automated method such as Akaike's information criterion (*AIC*) to determine the autoregressive order offers obvious practical advantages, again in contrast with the situation that prevails for the block bootstrap, whereby generic selection rules for the block length are unavailable. These deficiencies identified in the block bootstrap technique are likely to be manifest with long range dependent data *a-fortiori*, suggesting that the sieve bootstrap is likely to be even more favoured for fractionally integrated processes. For a review of block and sieve bootstrap methods and further discussion of their associated features see [Politis \(2003\)](#).

We illustrate our proposed methods by means of a simulation study, in which we examine the sieve bootstrap approximation to the sampling distribution of two types of statistic that satisfy the relevant conditions for the convergence results to hold. Firstly, we compare and contrast the performance of the raw and the pre-filtered sieve bootstrap in correctly characterizing the known finite sample properties of the sample mean under long memory. In particular, we investigate the previously noted tendency of bootstrap techniques to underestimate the true variance of the sample mean in this setting ([Hesterberg, 1997](#)). The pre-filtering is shown to correct for the distinct underestimation of the sampling variance still produced by the raw sieve, the higher-order accuracy of the latter notwithstanding. Secondly, we document the performance of the two bootstrap methods in estimating the (unknown) sampling distributions of selected autocorrelation coefficients. We undertake two exercises here. We begin by comparing the estimates of the sampling distributions produced by the (raw) sieve bootstrap with those produced via an Edgeworth approximation, in the region of the

parameter space where such an approximation is valid (see [Lieberman et al., 2001](#)). The bootstrap method is shown to produce distributions that are visually indistinguishable from those produced by the second-order Edgeworth expansion which, in turn, replicate the Monte Carlo estimates. Encouraged by the accuracy of the bootstrap method in the case in which an analytical finite sample comparator is available, we then proceed to assess the relative performance of the two alternative sieve bootstrap methods - raw and pre-filtered - in the part of the parameter space in which it is not. The pre-filtered method (in particular) is shown to produce particularly accurate estimates of the “true” (Monte Carlo) distributions in this region, auguring well for its general usefulness in empirical settings.

The paper proceeds as follows. Section 2 briefly outlines the statistical properties of autoregressive approximations to fractionally integrated processes, and summarizes the properties of the raw sieve bootstrap in this context. In Section 3 we present relevant Edgeworth expansions for a given class of statistics, and exploit these representations to establish the stated error rates for the raw sieve bootstrap technique. Section 4 outlines the methodology underlying the pre-filtered sieve bootstrap and presents the associated theory indicating the improvement obtained thereby. Details of the simulation study are given in Section 5. Section 6 closes the paper with some concluding remarks.

2 Long memory processes, autoregressive approximation, and the sieve bootstrap

Let $y(t)$ for $t \in \mathbb{Z}$ denote a linearly regular, covariance-stationary process with representation as in (1.1) where the innovations and the impulse response coefficients satisfy the following conditions:

Assumption 1 *The innovation process $\varepsilon(t)$ is ergodic and,*

$$E[\varepsilon(t) \mid \mathcal{E}_{t-1}] = 0 \quad \text{and} \quad E[\varepsilon(t)^2 \mid \mathcal{E}_{t-1}] = \sigma^2, \quad (\text{ass1})$$

where \mathcal{E}_t denotes the σ -algebra of events determined by $\varepsilon(s)$, $s \leq t$. Furthermore, $E[\varepsilon(t)^4] < \infty$.

Assumption 2 *The transfer function $k(z) = \sum_{j \geq 0} k(j)z^j$ in the representation of the process $y(t)$ is given by $k(z) = \kappa(z)/(1-z)^d$ where $|d| < 0.5$ and $\kappa(z)$ satisfies $\kappa(z) \neq 0$, $|z| \leq 1$, and $\sum_{j \geq 0} j|\kappa(j)| < \infty$.*

Assumption 1 imposes a classical martingale difference structure on the innovations, the critical property of such a process that drives the asymptotic results being that

a martingale difference is uncorrelated with any measurable function of its own past. Assumption 2 rules out the possibility of a root at unity in $\kappa(z)$ canceling with $(1 - z)^d$ and implies that the underlying process admits an infinite-order autoregressive ($AR(\infty)$) representation. Assumptions 1 and 2 incorporate quite a wide class of linear processes, including the popular ARFIMA family of models introduced by Granger and Joyeux (1980) and Hosking (1980).

Under Assumptions 1 and 2 $y(t) = \bar{y}(t) + \varepsilon(t)$ where the linear predictor

$$\bar{y}(t) = \sum_{j=1}^{\infty} \pi(j)y(t-j), \quad \sum_{j=1}^{\infty} \pi(j)z^j = 1 - k(z)^{-1},$$

is the minimum mean squared error predictor (MMSEP) of $y(t)$ based on the infinite past. The MMSEP of $y(t)$ based only on the finite past is then

$$\bar{y}_h(t) = \sum_{j=1}^h \pi_h(j)y(t-j) \equiv - \sum_{j=1}^h \phi_h(j)y(t-j), \quad (2.1)$$

where the minor reparameterization from π_h to ϕ_h allows us, on also defining $\phi_h(0) = 1$, to conveniently write the corresponding prediction error as

$$\varepsilon_h(t) = \sum_{j=0}^h \phi_h(j)y(t-j). \quad (2.2)$$

The finite-order autoregressive coefficients $\phi_h(1), \dots, \phi_h(h)$ can be deduced from the Yule-Walker equations

$$\sum_{j=0}^h \phi_h(j)\gamma(j-k) = \delta_0(k)\sigma_h^2, \quad k = 0, 1, \dots, h, \quad (2.3)$$

in which $\gamma(\tau) = \gamma(-\tau) = E[y(t)y(t-\tau)]$, $\tau = 0, 1, \dots$ is the autocovariance function of the process $y(t)$, $\delta_0(k)$ is Kronecker's delta (i.e., $\delta_0(k) = 0 \forall k \neq 0$; $\delta_0(0) = 1$), and

$$\sigma_h^2 = E[\varepsilon_h(t)^2] \quad (2.4)$$

is the prediction error variance associated with $\bar{y}_h(t)$.

The use of finite-order autoregressive models to approximate an unknown (but suitably regular) process therefore requires that the optimal predictor $\bar{y}_h(t)$ determined from the autoregressive model of order h ($AR(h)$) be a good approximation to the “infinite-order” predictor $\bar{y}(t)$ for sufficiently large h . The asymptotic validity and properties of $AR(h)$ models when $h \rightarrow \infty$ with the sample size T under regularity conditions that admit non-summable processes were established in Poskitt (2007). Briefly, the order- h prediction error $\varepsilon_h(t)$ converges to $\varepsilon(t)$ in mean-square, the estimated sample-based covariances converge to their population counterparts, though at

a slower rate than for a conventionally stationary process, and the least squares and Yule-Walker estimators of the coefficients of the $AR(h)$ approximation are asymptotically equivalent and consistent. Furthermore, order selection by AIC is asymptotically efficient in the sense of being equivalent to minimizing Shibata's (1980) figure of merit, discussed in more detail in Section 5 in the context of the simulation experiment reported therein. The sieve bootstrap, which works by "whitening" the data using an $AR(h)$ approximation, with the dynamics of the process captured in the fitted autoregression, is accordingly a plausible semi-parametric bootstrap technique for long memory processes. Details of its application to fractional processes are given in Poskitt (2008).

For convenience we present here the basic steps needed to generate a sieve bootstrap realization of a process $y(t)$ (referred to as the sieve bootstrap (SBS) algorithm hereafter):

SB1. Given data $y(t)$, $t = 1, \dots, T$, calculate the parameter estimates of the $AR(h)$ approximation, denoted by $\bar{\phi}_h(1), \dots, \bar{\phi}_h(h)$ and $\bar{\sigma}_h^2$, and evaluate the residuals,

$$\bar{\varepsilon}_h(t) = \sum_{j=0}^h \bar{\phi}_h(j) y(t-j), \quad t = 1, \dots, T,$$

using $y(1-j) = y(T-j+1)$, $j = 1, \dots, h$, as initial values. From $\bar{\varepsilon}_h(t)$, $t = 1, \dots, T$, construct the standardized residuals $\tilde{\varepsilon}_h(t) = (\bar{\varepsilon}_h(t) - \bar{\varepsilon}_h) / s_{\bar{\varepsilon}_h}$, where $\bar{\varepsilon}_h = T^{-1} \sum_{t=1}^T \bar{\varepsilon}_h(t)$ and $s_{\bar{\varepsilon}_h}^2 = T^{-1} \sum_{t=1}^T (\bar{\varepsilon}_h(t) - \bar{\varepsilon}_h)^2$.

SB2. Let $\varepsilon_h^+(t)$, $t = 1, \dots, T$, denote a simple random sample of *i.i.d.* values drawn from

$$U_{\bar{\varepsilon}_h, T}(e) = T^{-1} \sum_{t=1}^T \mathbf{1}\{\tilde{\varepsilon}_h(t) \leq e\},$$

the probability distribution function that places a probability mass of $1/T$ at each of $\tilde{\varepsilon}_h(t)$, $t = 1, \dots, T$. Set $\varepsilon_h^*(t) = \bar{\sigma}_h \varepsilon_h^+(t)$, $t = 1, \dots, T$.

SB3. Construct the sieve bootstrap realization $y^*(1), \dots, y^*(T)$ where $y^*(t)$ is generated from the autoregressive process

$$\sum_{j=0}^h \bar{\phi}_h(j) y^*(t-j) = \varepsilon_h^*(t), \quad t = 1, \dots, T,$$

initiated at $y^*(1-j) = y(\tau-j+1)$, $j = 1, \dots, h$, where τ has the discrete uniform distribution on the integers h, \dots, T .

Crucially, in the fractional case the rate of convergence of the coefficient estimates $\bar{\phi}_h(1), \dots, \bar{\phi}_h(h)$ evaluated in Step SB1 is dependent upon the value of the fractional index d .

Theorem 3 Let $\sum_{j=0}^h \bar{\phi}_h(j) z^j$ denote the Burg, least squares or Yule-Walker estimator of $\sum_{j=0}^h \phi_h(j) z^j$. If $y(t)$ is a stationary process that satisfies Assumptions 1 and 2 (given below) then for all $h \leq H_T = a(\log T)^c$, $a > 0$, $c < \infty$,

$$\sum_{j=1}^h |\bar{\phi}_h(j) - \phi_h(j)|^2 = O \left\{ h \left(\frac{\log T}{T} \right)^{1-2d'} \right\} \quad a.s.$$

where $d' = \max\{0, d\}$.

Proof: For the least squares and Yule-Walker estimators see [Poskitt \(2007, Theorem 5 and Corollary 1\)](#) and the associated discussion. For the Burg estimator the result then follows from [Poskitt \(1994, Theorem 1\)](#). ■

Now consider a statistic $\mathbf{s}_T = (s_{1T}, \dots, s_{mT})'$, where $s_{iT} = s_i(y(1), \dots, y(T))$ and each $s_i(\cdot)$ for $i = 1, \dots, m$ is a suitably smooth function of the time series values $y(1), \dots, y(T)$. Let $F_{\mathbf{s}_T}(\mathbf{s})$ denote the distribution function of \mathbf{s}_T under $(\Omega, \mathfrak{F}, P)$, the original probability space. Let \mathbf{s}_T^* be defined as for \mathbf{s}_T but with the observed realization replaced by $y^*(1), \dots, y^*(T)$, a realization obtained from the SBS algorithm, so that $\mathbf{s}_T^* = (s_{1T}^*, \dots, s_{mT}^*)'$ where $s_{iT}^* = s_i(y^*(1), \dots, y^*(T))$. Let $F_{\mathbf{s}_T^*}(\mathbf{s})$ denote the distribution of \mathbf{s}_T^* under $(\Omega^*, \mathfrak{F}^*, P^*)$, the bootstrap probability space. As with $F_{\mathbf{s}_T}(\mathbf{s})$, the analytical determination of $F_{\mathbf{s}_T^*}(\mathbf{s})$ is generally intractable, but by simulating a large number, B , of independent bootstrap realizations and calculating $\mathbf{s}_{T,b}^*$ for $b = 1, \dots, B$, we can approximate $F_{\mathbf{s}_T^*}(\mathbf{s})$ by the bootstrap empirical distribution function

$$\bar{F}_{\mathbf{s}_T^*, B}(\mathbf{s}) = B^{-1} \sum_{b=1}^B \mathbf{1}\{\mathbf{s}_{T,b}^* \leq \mathbf{s}\}. \quad (2.5)$$

By the (strong) Glivenko-Cantelli Theorem

$$\limsup_{B \rightarrow \infty} \sqrt{\frac{B}{2 \log \log B}} \sup_{\mathbf{s}} |\bar{F}_{\mathbf{s}_T^*, B}(\mathbf{s}) - F_{\mathbf{s}_T^*}(\mathbf{s})| \leq \frac{1}{2} \quad a.s.$$

and we can approximate $F_{\mathbf{s}_T^*}(\mathbf{s})$ arbitrarily closely by taking the number of bootstrap realizations sufficiently large. The idea behind the bootstrap is that the distribution of \mathbf{s}_T^* under $(\Omega^*, \mathfrak{F}^*, P^*)$ should mimic that of \mathbf{s}_T under $(\Omega, \mathfrak{F}, P)$ and we can therefore anticipate that $\bar{F}_{\mathbf{s}_T^*, B}(\mathbf{s})$ will also approximate $F_{\mathbf{s}_T}(\mathbf{s})$ closely provided $F_{\mathbf{s}_T^*}(\mathbf{s})$ is sufficiently near to $F_{\mathbf{s}_T}(\mathbf{s})$.

That the autoregressive sieve bootstrap provides a valid approximation to $F_{\mathbf{s}_T}(\mathbf{s})$ under the current assumptions can be established by generalizing the arguments of [Kreiss, Paparoditis and Politis \(2011\)](#) using the extension of Baxter's inequality due to [Inoue and Kasahara \(2006\)](#). It can be shown ([Poskitt, 2008](#)) that for the class of linear statistics considered in [Künsch \(1989, Section 2.1\)](#) and [Bühlmann \(1997, Section 3.3\)](#) we have $\eta(F_{\mathbf{s}_T^*}, F_{\mathbf{s}_T}) = o(T^{-1/2(1-2d')+\beta})$ for all $\beta > 0$, wherein $d' = \max\{0, d\}$ and

$\eta(F_X, F_Y)$ denotes Mallows' measure of the distance between two probability distributions F_X and F_Y . Mallows metric is equivalent to weak convergence (Bickel and Freedman, 1981, Lemma 8.3) and in conjunction with a convergence rate of $T^{-1/2(1-2d')+\beta}$ this intimates that use of the sieve bootstrap may be little better than applying a central limit approximation. However, in what follows we show that for a more restricted range of statistics (albeit one that intersects with the linear class) the convergence rate can be improved upon, and that the rate established by Choi and Hall (2000) in the short memory case can in fact be generalized to long memory processes.

3 Higher Order Improvements for the Sieve Bootstrap

Let us suppose that $F_{\mathbf{s}_T}(\mathbf{s})$ is absolutely continuous with respect to Lebesgue measure, differentiable for all \mathbf{s} , and that the following assumptions are satisfied.

Assumption 4 *There exists a function M_T (possibly stochastic) and a constant $M < \infty$ such that*

$$\|\mathbf{s}_T^* - \mathbf{s}_T\|^2 \leq mM_T T^{-1} \sum_{t=1}^T (y^*(t) - y(t))^2,$$

where M_T is bounded (in probability) by M .

Whilst defining a more restrictive class (overall) than the linear class, it remains the case that a broad range of statistics used in the analysis of time series satisfy Assumption 4, see (Poskitt, 2008, Lemma 1). As highlighted in the latter, this set includes the sample mean and the sample autocovariances, autocorrelations and partial autocorrelations. Further examples include the discrete Fourier transform and OLS regression coefficients. The former follows on setting

$$\mathbf{s}_T^* - \mathbf{s}_T = \frac{1}{(2\pi T)^{\frac{1}{2}}} \sum_{t=1}^T \begin{bmatrix} y^*(t) - y(t) \\ (y^*(t) - y(t)) \exp(-i2\pi t/T) \\ \vdots \\ (y^*(t) - y(t)) \exp(-i2\pi(T-1)t/T) \end{bmatrix},$$

with the validity of Assumption 4, using $M = 1/(2\pi)$, now a direct consequence of Parseval's theorem. For the latter, let $\mathbf{x}(t) = (x_1(t), \dots, x_k(t))'$ denote a vector of regressors that satisfy $\liminf_{T \rightarrow \infty} \lambda_{\min}[T^{-1} \sum_{t=1}^T \mathbf{x}(t)\mathbf{x}(t)'] \geq \lambda > 0$ and set $\mathbf{k}(t) = (\sum_{t=1}^T \mathbf{x}(t)\mathbf{x}(t)')^{-1} \mathbf{x}(t)$. Then

$$\|\mathbf{s}_T^* - \mathbf{s}_T\|^2 = \left\| \sum_{t=1}^T \mathbf{k}(t)(y^*(t) - y(t)) \right\|^2 \leq \sum_{t=1}^T \|\mathbf{k}(t)\|^2 \sum_{t=1}^T (y^*(t) - y(t))^2$$

and Assumption 4 holds because $\sum_{t=1}^T \|\mathbf{k}(t)\|^2 \leq k/\lambda T$. As a point of interest, applying the immediately preceding regression inequality to the log-periodogram regression estimator of d (Geweke and Porter-Hudak, 1983) and using the inequalities $\log(1+x) \leq x$, $x \geq 0$, and $\|z\| - \|z^*\| \leq |z - z^*|$ for any pair of complex numbers z and z^* , we also find that the log-periodogram regression estimator satisfies Assumption 4 on application of the bound on the discrete Fourier transform.

The following assumption implicitly characterizes moment conditions under which a valid Edgeworth expansion exists for statistics in the class described by Assumption 4.

Assumption 5 Let $\psi_T(\boldsymbol{\tau}) = E[\exp(i\boldsymbol{\tau}'\mathbf{s}_T)]$ denote the characteristic function of \mathbf{s}_T where $\boldsymbol{\tau} = (\tau_1, \dots, \tau_m)'$ and let $\partial^j \log \psi_T(\boldsymbol{\tau})/\partial \boldsymbol{\tau}^j$ denote the vector of j th-order partial derivatives corresponding to $\partial^j \log \psi_T(\boldsymbol{\tau})/\partial \tau_1^{j_1} \dots \partial \tau_m^{j_m}$ for all non-negative integers j_1, \dots, j_m satisfying $\sum_{l=1}^m j_l = j$. Then firstly, for any $\delta > 0$ and some integer $r \geq 3$ the conditions

$$\int_{\|\boldsymbol{\tau}\| > \delta\sqrt{T}} |\psi_T(\boldsymbol{\tau})|^2 d\boldsymbol{\tau} = o(T^{2-r}) \quad \text{and} \quad \int_{\|\boldsymbol{\tau}\| > \delta\sqrt{T}} \left| \frac{\partial^s \psi_T(\boldsymbol{\tau})}{\partial \tau_l^s} \right|^2 d\boldsymbol{\tau} = O(T^{1-r}), \quad l = 1, \dots, m,$$

hold where $s = [m/2] + 1$. Secondly, $\partial^q \log \psi_T(\boldsymbol{\tau})/\partial \boldsymbol{\tau}^q$ exists for all $\boldsymbol{\tau}$ in a neighbourhood of the origin and $\lim_{\|\boldsymbol{\tau}\| \rightarrow 0} T^{-1} \partial^q \log \psi_T(\boldsymbol{\tau})/\partial \boldsymbol{\tau}^q$ exists as $T \rightarrow \infty$ for all $q = 1, \dots, q' = \max\{s, r + 1\}$.

Here E denotes expectation taken with respect to the probability measure induced by the original probability space $(\Omega, \mathfrak{F}, P)$. Assumption 5 summarizes Assumptions 1 and 2 of Taniguchi (1984), which in turn encompass Assumptions 2 through 4 of Durbin (1980), to which we refer for an in depth discussion. In any particular instance, satisfaction of the conditions in Assumption 5 must be ascertained and may occur only in particular parts of the parameter space, such as in the case of the sample autocorrelation function investigated in Section 5.3 (See the Appendix for details).

Let $\mathbf{V}_T = T^{-1} E[(\mathbf{s}_T - E[\mathbf{s}_T])(\mathbf{s}_T - E[\mathbf{s}_T])']$ and set $\boldsymbol{\zeta}_T = \mathbf{V}_T^{-1/2} T^{-\frac{1}{2}}(\mathbf{s}_T - E[\mathbf{s}_T])$. If we suppose that $\mathbf{V}_T = \mathbf{V} + o(1)$ where \mathbf{V} is positive definite, then Assumption 5 ensures the validity of the Edgeworth approximation

$$P(\boldsymbol{\zeta}_T \leq \mathbf{z}) = G(\mathbf{z}) + \sum_{j=3}^r T^{1-j/2} p_j(\mathbf{z}, \mathbf{K}_r) g(\mathbf{z}) + o(T^{1-r/2}) \quad (3.1)$$

uniformly in \mathbf{z} , where $G(\mathbf{z})$ denotes the distribution function of a Gaussian $\mathbb{N}(\mathbf{0}, \mathbf{I}_m)$ random vector, $g(\mathbf{z})$ the corresponding density, and $p_j(\mathbf{z}, \mathbf{K}_r)$ is a polynomial function of degree j in \mathbf{z} whose coefficients are polynomials in the elements of the cumulants $\mathbf{K}_r = (\mathbf{k}'_1, \dots, \mathbf{k}'_r)'$, $\mathbf{k}_r = i^{-r} \partial^r \log \psi_T(\mathbf{0})/\partial \boldsymbol{\tau}^r$. See Theorem 1 of Taniguchi (1984) and

Durbin (1980).¹

Similarly, if E^* denotes expectation taken with respect to the probability space $(\Omega^*, \mathfrak{F}^*, P^*)$ and $\zeta_T^* = \mathbf{V}_T^{*-1/2} T^{-\frac{1}{2}} (\mathbf{s}_T^* - E^*[\mathbf{s}_T^*])$ where $\mathbf{V}_T^* = T^{-1} E^* [(\mathbf{s}_T^* - E^*[\mathbf{s}_T^*])(\mathbf{s}_T^* - E^*[\mathbf{s}_T^*])']$, then under appropriate regularity

$$P^*(\zeta_T^* \leq \mathbf{z}) = G(\mathbf{z}) + \sum_{j=3}^r T^{1-j/2} p_j(\mathbf{z}, \mathbf{K}_r^*) g(\mathbf{z}) + o(T^{1-r/2}), \quad (3.2)$$

where $\mathbf{K}_r^* = (\mathbf{k}_1^*, \dots, \mathbf{k}_r^*)'$, $\mathbf{k}_r^* = \iota^{-r} \{\partial^r \log \psi_T^*(\mathbf{0}) / \partial \boldsymbol{\tau}^r\}$, $\psi_T^*(\boldsymbol{\tau}) = E^*[\exp(\iota \boldsymbol{\tau}' \mathbf{s}_T^*)]$.

A comparison of (3.1) and (3.2) for $r \geq 4$ now indicates that

$$\sup_{\mathbf{z}} |P^*(\zeta_T^* \leq \mathbf{z}) - P(\zeta_T \leq \mathbf{z})| = T^{-1/2} O(\|\mathbf{K}_r^* - \mathbf{K}_r\|) + o(T^{-1}). \quad (3.3)$$

Noting that P^* depends on P so the elements of \mathbf{K}_r^* , which are constants relative to P^* , are random variables relative to P , we see that if $\|\mathbf{K}_4^* - \mathbf{K}_4\| = o_p(T^{-1/2} \varrho_T)$ then (3.3) implies that the bootstrap probability will have an error rate of $O_p(T^{-1} \varrho_T)$. In their investigation of coverage accuracy Choi and Hall (2000, Appendix A.2) used this type of argument when analyzing the subset of linear statistics characterized by polynomial products; and it was also employed by Andrews et al. (2006) to show that the parametric bootstrap based on the (approximate) MLE of parameters in a Gaussian long memory model achieves an error rate of order $T^{-1} \log T$ for a one sided confidence interval. Using this approach we can establish analogous results for the sieve bootstrap in the long memory case, and for the class of statistics encompassed by Assumption 4.

Theorem 6 *Suppose that the statistic \mathbf{s}_T satisfies Assumption 4 and Assumption 5 with $r \geq 4$ when calculated from a process $y(t)$ that also satisfies Assumptions 1 and 2. Then for all $\beta > 0$*

$$\sup_{\mathbf{z}} |P^*(\zeta_T^* \leq \mathbf{z}) - P(\zeta_T \leq \mathbf{z})| = O_p(T^{-(1-d')+\beta}).$$

The proof of Theorem 6 relies on the following lemma. The heuristics behind the proof are straightforward; convergence of Mallow's metric implies convergence in distribution and hence, via the Cramér-Levy continuity theorem, convergence of the characteristic function and the associated moments and cumulants (See Lemma 8.3 of Bickel and Freedman, 1981).

Lemma 7 *Suppose that the process $y(t)$ satisfies Assumptions 1 and 2, and that the statistic \mathbf{s}_T satisfies Assumption 4. Then $E[E^*[\|\mathbf{s}_T^* - \mathbf{s}_T\|^2]] = o(T^{-(1-2d')+\beta})$ for all*

¹ We have thus far supposed that \mathbf{s}_T is a continuous random variable. For extension to the lattice case see Durbin (1980, §5.4)

$\beta > 0$. Moreover, for any $\epsilon > 0$ and $\eta > 0$, no matter how small, there exists a $T_{\epsilon, \eta}$ such that

$$P\left(T^{-1/2(1-2d')+\beta} |\psi_T^*(\boldsymbol{\tau}) - \psi_T(\boldsymbol{\tau})| < \epsilon\right) > 1 - \eta$$

for all $T > T_{\epsilon, \eta}$ uniformly in $\boldsymbol{\tau}$, $\|\boldsymbol{\tau}\| \leq T^{\beta/2}$.

Proof: Given that \mathbf{s}_T satisfies Assumption 4, it follows that $\|\mathbf{s}_T^* - \mathbf{s}_T\|^2 \leq mMT^{-1} \sum_{t=1}^T (y^*(t) - y(t))^2$ where $M < \infty$. Arguing as in the proof of Theorem 2 of (Poskitt, 2008, p.246-248) we therefore have $E[E^*[\|\mathbf{s}_T^* - \mathbf{s}_T\|^2]] \leq mME[E^*[(y^*(t) - y(t))^2]] = o(T^{-(1-2d')+\beta})$ for all $\beta > 0$, which yields the first part of the lemma.

To prove the second part of the lemma note that since $\exp(ix)$ is continuous with a continuous and uniformly bounded derivative it satisfies a Lipschitz condition. Thus, for all $\boldsymbol{\tau}$ such that $\|\boldsymbol{\tau}\| \leq T^{\beta/2}$ there exists a $K < \infty$ such that $|\exp(i\boldsymbol{\tau}'\mathbf{x}) - \exp(i\boldsymbol{\tau}'\mathbf{y})| \leq KT^{\beta/2}\|\mathbf{x} - \mathbf{y}\|$. Then, as in Bickel and Freedman (1981, p. 1212),

$$E[|\psi_T^*(\boldsymbol{\tau}) - \psi_T(\boldsymbol{\tau})|] \leq E[E^*[\|\exp(i\boldsymbol{\tau}\mathbf{s}_T^*) - \exp(i\boldsymbol{\tau}\mathbf{s}_T)\|]] \leq KT^{\beta/2}E[E^*[\|\mathbf{s}_T^* - \mathbf{s}_T\|]].$$

But $E[E^*[\|\mathbf{s}_T^* - \mathbf{s}_T\|]] \leq E[E^*[\|\mathbf{s}_T^* - \mathbf{s}_T\|^2]]^{1/2} = o(T^{-1/2(1-2d')-\beta})$. Application of Markov's inequality completes the proof. \blacksquare

Corollary 8 Suppose that the process $y(t)$ satisfies Assumptions 1 and 2, and that the statistic \mathbf{s}_T satisfies Assumptions 4 and 5. Then for all $\beta > 0$ we have $\|\mathbf{K}_q^* - \mathbf{K}_q\| = o_p(T^{-1/2+d'+\beta})$, $q = 1, \dots, q' = \max\{[m/2] + 1, r + 1\}$.

Proof: Using the expression $\log \psi_T^*(\boldsymbol{\tau}) - \log \psi_T(\boldsymbol{\tau}) = \log(1 + (\psi_T^*(\boldsymbol{\tau}) - \psi_T(\boldsymbol{\tau}))/\psi_T(\boldsymbol{\tau}))$ and the fact that $\log(1+x) = x + O(|x|^2)$ for x in a neighbourhood of the origin we have $\log \psi_T^*(\boldsymbol{\tau}) - \log \psi_T(\boldsymbol{\tau}) = (\psi_T^*(\boldsymbol{\tau}) - \psi_T(\boldsymbol{\tau}))/\psi_T(\boldsymbol{\tau}) + O(|(\psi_T^*(\boldsymbol{\tau}) - \psi_T(\boldsymbol{\tau}))/\psi_T(\boldsymbol{\tau})|^2) = o_p(T^{-1/2(1-2d')+\beta})$ uniformly in $\boldsymbol{\tau}$ by Lemma 7.

Now set

$$\begin{aligned} \varphi_T^*(\mathbf{t}; \boldsymbol{\tau}) &= \frac{\log \psi_T^*(\mathbf{t}) - \log \psi_T^*(\boldsymbol{\tau})}{\|\mathbf{t} - \boldsymbol{\tau}\|} - \left(\frac{\partial \log \psi_T(\boldsymbol{\tau})}{\partial \boldsymbol{\tau}} \right)' \frac{\mathbf{t} - \boldsymbol{\tau}}{\|\mathbf{t} - \boldsymbol{\tau}\|} \quad \text{and} \\ \Delta_T^*(\mathbf{t}; \boldsymbol{\tau}) &= \frac{\log \psi_T^*(\mathbf{t}) - \log \psi_T^*(\boldsymbol{\tau})}{\|\mathbf{t} - \boldsymbol{\tau}\|} - \frac{\log \psi_T(\mathbf{t}) - \log \psi_T(\boldsymbol{\tau})}{\|\mathbf{t} - \boldsymbol{\tau}\|}, \end{aligned}$$

for $\mathbf{t} \neq \boldsymbol{\tau}$, and let $\epsilon > 0$ be given. Then

$$\begin{aligned} |\varphi_T^*(\mathbf{t}; \boldsymbol{\tau})| &\leq |\Delta_T^*(\mathbf{t}; \boldsymbol{\tau})| \\ &\quad + \left| \frac{\log \psi_T(\mathbf{t}) - \log \psi_T(\boldsymbol{\tau})}{\|\mathbf{t} - \boldsymbol{\tau}\|} - \left(\frac{\partial \log \psi_T(\boldsymbol{\tau})}{\partial \boldsymbol{\tau}} \right)' \frac{\mathbf{t} - \boldsymbol{\tau}}{\|\mathbf{t} - \boldsymbol{\tau}\|} \right| \end{aligned}$$

and by definition of the differential (Apostol, 1960, Section 6.4)

$$\lim_{\|\mathbf{t} - \boldsymbol{\tau}\| \rightarrow 0} |\varphi_T^*(\mathbf{t}; \boldsymbol{\tau})| \leq \lim_{\|\mathbf{t} - \boldsymbol{\tau}\| \rightarrow 0} |\Delta_T^*(\mathbf{t}; \boldsymbol{\tau})| + \epsilon.$$

Since $\log \psi_T^*(\boldsymbol{\tau}) - \log \psi_T(\boldsymbol{\tau}) = o_p(T^{-1/2(1-2d')+\beta})$ uniformly in $\boldsymbol{\tau}$ we can interchange limiting operations (Apostol, 1960, Theorem 13.3) to give

$$\begin{aligned} \lim_{T \rightarrow \infty} \lim_{\|\mathbf{t}-\boldsymbol{\tau}\| \rightarrow 0} |\Delta_T^*(\mathbf{t}; \boldsymbol{\tau})| &\leq \lim_{\|\mathbf{t}-\boldsymbol{\tau}\| \rightarrow 0} \lim_{T \rightarrow \infty} \frac{|\log \psi_T^*(\mathbf{t}) - \log \psi_T(\mathbf{t})| + |\log \psi_T^*(\boldsymbol{\tau}) - \log \psi_T(\boldsymbol{\tau})|}{\|\mathbf{t} - \boldsymbol{\tau}\|} \\ &= o_p(T^{-1/2(1-2d')+\beta}). \end{aligned}$$

Hence we can conclude that for all T sufficiently large $\lim_{\|\mathbf{t}-\boldsymbol{\tau}\| \rightarrow 0} |\varphi_T^*(\mathbf{t}; \boldsymbol{\tau})| \leq 2\epsilon$ and $\log \psi_T^*(\boldsymbol{\tau})$ has a differential at $\boldsymbol{\tau}$, since ϵ is arbitrary, and

$$\lim_{T \rightarrow \infty} \left| \frac{\partial \log \psi_T^*(\boldsymbol{\tau})}{\partial \tau_j} - \frac{\partial \log \psi_T(\boldsymbol{\tau})}{\partial \tau_j} \right| \leq \lim_{h \rightarrow 0} \lim_{T \rightarrow \infty} |\Delta_T^*(\boldsymbol{\tau} + h\mathbf{u}_j; \boldsymbol{\tau})| = o_p(T^{-1/2(1-2d')+\beta}),$$

where $\mathbf{u}_j = (0, \dots, 0, 1, 0, \dots, 0)'$, the j th unit vector, the existence of the gradient vector $\partial \log \psi_T^*(\boldsymbol{\tau})/\partial \boldsymbol{\tau}$ being part of the conclusion (See Apostol, 1960, Theorem 6.13). Thus, by definition of the first-order cumulant, we have $\|\mathbf{K}_1^* - \mathbf{K}_1\| = o_p(T^{-1/2+d'+\beta})$.

A parallel argument, with $\log \psi_T^*(\cdot)$ and $\log \psi_T(\cdot)$ replaced by $\partial^j \log \psi_T^*(\cdot)/\partial \boldsymbol{\tau}^j$ and $\partial^j \log \psi_T(\cdot)/\partial \boldsymbol{\tau}^j$, respectively, and $\partial \log \psi_T(\cdot)/\partial \boldsymbol{\tau}$ replaced by $\partial^{j+1} \log \psi_T(\cdot)/\partial \boldsymbol{\tau}^{j+1}$, shows that $\partial^{j+1} \log \psi_T^*(\cdot)/\partial \boldsymbol{\tau}^{j+1}$ exists and $\|\mathbf{K}_{j+1}^* - \mathbf{K}_{j+1}\| = o_p(T^{-1/2+d'+\beta})$. Induction on $j = 1, \dots, q'$ completes the proof. \blacksquare

Proof of Theorem 6: By construction the bootstrap innovations $\varepsilon_h^*(t)$ in Step SB2 of the sieve bootstrap satisfy Assumption 1, and the sieve bootstrap process $y^*(t)$ produced in Step SB3 satisfies Assumption 2. By definition, the statistics \mathbf{s}_T and \mathbf{s}_T^* satisfy Assumption 4 and Assumption 5 with $r \geq 4$, and Assumption 5 validates the formal Edgeworth expansions in (3.2) and (3.1). Corollary 8 (using Lemma 7) implies that $\|\mathbf{K}_4^* - \mathbf{K}_4\| = o_p(T^{-1/2}\varrho_T)$ where $\varrho_T = o(T^{d'+\beta})$ and Theorem 6 then follows from equation (3.3). \blacksquare

Theorem 6 indicates the refinements that are possible using the sieve bootstrap. For example, $S(q) = \{\mathbf{z} : \mathbf{z}'\mathbf{z} \leq q\}$ is a compact, Borel-measurable set in \mathbb{R}^m that has finite probability measure with respect to both P and P^* . Now let q_α^* be such that the Lebesgue–Stieltjes integral satisfies the following equality,

$$\int_{S(q_\alpha^*)} dP^*(\boldsymbol{\zeta}_T^* \leq \mathbf{z}) = 1 - \alpha.$$

Then $S(q_\alpha^*)$ is a raw sieve bootstrap $(1 - \alpha)100\%$ elliptical percentile set for \mathbf{s}_T . Now, from Theorem 6 it follows that

$$\begin{aligned} |(1 - \alpha) - \int_{S(q_\alpha^*)} dP(\boldsymbol{\zeta}_T \leq \mathbf{z})| &\leq \int_{S(q_\alpha^*)} |dP^*(\boldsymbol{\zeta}_T^* \leq \mathbf{z}) - dP(\boldsymbol{\zeta}_T \leq \mathbf{z})| \\ &= \frac{(\pi q_\alpha^*)^{m/2}}{\Gamma(m/2 + 1)} O_p(T^{-(1-d')+\beta}) \end{aligned}$$

for all $\beta > 0$. This leads to a coverage probability for $S(q_\alpha^*)$ of $(1 - \alpha) + O_p(T^{-(1-d)+\beta})$ for all $\beta > 0$ when $d \in (0, 0.5)$, the long memory case, compared to $(1 - \alpha) + O_p(T^{-1+\beta})$ when $d \in (-0.5, 0]$, the short memory and anti-persistent cases (*cf.* Choi and Hall, 2000). Calibration of the percentile sets using the double-bootstrap may be possible, but we will not pursue this here. We will, however, investigate in the following section an adaptation of the sieve bootstrap that improves the convergence rate by removing the dependence on the fractional index d .

4 The Pre-Filtered Sieve Bootstrap

Theorem 3 indicates that the convergence of $\sum_{j=0}^h \bar{\phi}_h(j) z^j$ to $\sum_{j=0}^h \phi_h(j) z^j$ is slower the larger is the value of d , and Theorem 6 shows that this feature is passed on to the raw sieve bootstrap itself (and the associated coverage probabilities of sets). Specifically, the closer is d to zero the closer the convergence rate will be to the rate achieved with short memory and anti-persistent processes, namely $O_p(T^{-1+\beta})$. Given the empirical regularity of estimated values of d in the range $(0, 0.5)$, calculating a preliminary estimate of d and constructing a filtered version of the data to which the AR approximation and sieve bootstrap are applied before inverse filtering, may therefore yield advantages in terms of convergence.

With this in mind, let us suppose that a preliminary estimate \hat{d} of d is available such that $\hat{d} - d \in N_\delta = \{x : |x| < \delta\}$ where $0 < \delta < 0.5$. For any $d > -1$ let $\alpha_j^{(d)}$, $j = 0, 1, 2, \dots$, denote the coefficients of the fractional difference operator when expressed in terms of its binomial expansion,

$$\begin{aligned} (1 - z)^d &= \sum_{j=0}^{\infty} \alpha_j^{(d)} z^j = 1 + \sum_{j=1}^{\infty} \left(\frac{\Gamma(j - d)}{\Gamma(-d)\Gamma(j + 1)} \right) z^j \\ &= 1 + \sum_{j=1}^{\infty} \left(\prod_{0 < k \leq j} \frac{k - 1 - d}{k} \right) z^j, \end{aligned}$$

and set

$$w(t) = \sum_{j=0}^{t-1} \alpha_j^{(d)} y(t - j), \quad t = 1, \dots, T.$$

Using the preliminary estimate \hat{d} , pre-filtered sieve bootstrap realizations of $y(t)$ can now be generated as follows:

PFSBS1. Calculate the coefficients of the filter $(1 - z)^{\hat{d}}$ and from the data generate the filtered values

$$\hat{w}(t) = \sum_{j=0}^{t-1} \alpha_j^{(\hat{d})} y(t - j)$$

for $t = 1, \dots, T$.

PFSBS2. Fit an AR approximation to $\widehat{w}(t)$ and generate a sieve bootstrap sample $\widehat{w}^*(t)$, $t = 1, \dots, T$, of the filtered data as in Steps SB1–SB3 of the SBS algorithm.

PFSBS3. Using the coefficients of the (inverse) filter $(1 - z)^{-\widehat{d}}$ construct a corresponding pre-filtered sieve bootstrap draw

$$\widehat{y}^*(t) = \sum_{j=0}^{t-1} \alpha_j^{(-\widehat{d})} \widehat{w}^*(t-j) \quad (4.1)$$

of $y(t)$ for $t = 1, \dots, T$.

We will refer to this as the PFSBS algorithm.

Note that the process

$$(1 - z)^{\widehat{d}} y(t) = \frac{\kappa(z)}{(1 - z)^{d - \widehat{d}}} \varepsilon(t)$$

has fractional index $d - \widehat{d}$. By assumption $|d - \widehat{d}| < \delta$ and the error in the AR approximation fitted in Step PFSBS2 will accordingly be of order $O(h(\log T/T)^{1-2\delta})$ or smaller (Theorem 3). That this level of accuracy is transferred to the pre-filtered sieve bootstrap realizations $\widehat{y}^*(t)$ of $y(t)$, via the sieve bootstrap draws $\widehat{w}^*(t)$ of $\widehat{w}(t)$, and hence to the pre-filtered sieve bootstrap approximation to the sampling distribution of the statistic \mathbf{s}_T , rests upon the following proposition.

Proposition 9 *Suppose that the process $y(t)$ satisfies Assumptions 1 and 2. Let \widehat{d} be such that $\widehat{d} \in (-0.5, 0.5(1 - \epsilon))$ for some $\epsilon > 0$. Then there exists a constant $G < \infty$, independent of \widehat{d} , such that $E[E^*[(\widehat{y}^*(t) - y(t))^2]] \leq GE[E^*[(\widehat{w}^*(t) - \widehat{w}(t))^2]]$.*

Proof: By construction

$$y(t) = \sum_{j=0}^{t-1} \alpha_j^{(-\widehat{d})} \widehat{w}(t-j),$$

and subtracting on the left and right hand sides in (4.1) it follows that

$$\widehat{y}^*(t) - y(t) = \sum_{j=0}^{t-1} \alpha_j^{(-\widehat{d})} \{\widehat{w}^*(t-j) - \widehat{w}(t-j)\} \quad t = 1, \dots, T \quad (4.2)$$

for all possible pairs $(y(t), \widehat{y}^*(t))$, $t = 1, \dots, T$, in the product space generated by $(\Omega \otimes \Omega^*, \mathfrak{F} \otimes \mathfrak{F}^*, P(P^*))$ with joint distribution corresponding to the marginal and conditional probability measures P and P^* .

Now let $Z_{\{\widehat{w}^* - \widehat{w}\}}^T(\lambda)$, $\lambda \in [0, 2\pi]$, denote the finite sample spectral measure associated with the process $\widehat{w}^*(t) - \widehat{w}(t)$, $t = 1, \dots, T$, which we define to be

$$Z_{\{\widehat{w}^* - \widehat{w}\}}^T(\lambda) = \frac{1}{2\pi} \sum_{t=1}^T \frac{e^{-i\lambda t} - 1}{it} \{\widehat{w}^*(t) - \widehat{w}(t)\} = \frac{1}{2\pi} \int_0^\lambda z_{\{\widehat{w}^* - \widehat{w}\}}^T(\omega) d\omega$$

where $z_{\{\widehat{w}^* - \widehat{w}\}}^T(\omega) = \sum_{t=1}^T \{\widehat{w}^*(t) - \widehat{w}(t)\} e^{-i\omega t}$. Then we have

$$\begin{aligned} \widehat{w}^*(t) - \widehat{w}(t) &= \frac{1}{2\pi} \int_0^{2\pi} e^{i\lambda t} dZ_{\{\widehat{w}^* - \widehat{w}\}}^T(\lambda) \\ &= \frac{1}{T} \sum_{s=0}^{T-1} e^{i\frac{2\pi s}{T}t} z_{\{\widehat{w}^* - \widehat{w}\}}^T(2\pi s/T), \quad t = 1, \dots, T, \end{aligned}$$

and direct substitution into equation (4.2) yields the equivalent representations

$$\begin{aligned} \widehat{y}^*(t) - y(t) &= \frac{1}{2\pi} \int_0^{2\pi} \sum_{j=0}^{t-1} \alpha_j^{(-\widehat{d})} e^{i\lambda(t-j)} dZ_{\{\widehat{w}^* - \widehat{w}\}}^T(\lambda) \\ &= \frac{1}{T} \sum_{s=0}^{T-1} \sum_{j=0}^{t-1} \alpha_j^{(-\widehat{d})} e^{i\frac{2\pi s}{T}(t-j)} z_{\{\widehat{w}^* - \widehat{w}\}}^T(2\pi s/T), \quad t = 1, \dots, T. \end{aligned}$$

The Cauchy-Schwartz inequality now gives us

$$\begin{aligned} |\widehat{y}^*(t) - y(t)| &= \left| \frac{1}{T} \sum_{s=0}^{T-1} \left(\sum_{j=0}^{t-1} \alpha_j^{(-\widehat{d})} e^{-i\frac{2\pi s}{T}j} \right) e^{i\frac{2\pi s}{T}t} z_{\{\widehat{w}^* - \widehat{w}\}}^T(2\pi s/T) \right| \\ &\leq \left(\frac{1}{T} \sum_{s=0}^{T-1} \left| \sum_{j=0}^{t-1} \alpha_j^{(-\widehat{d})} e^{-i\frac{2\pi s}{T}j} \right|^2 \right)^{1/2} \cdot \left(\frac{1}{T} \sum_{s=0}^{T-1} |z_{\{\widehat{w}^* - \widehat{w}\}}^T(2\pi s/T)|^2 \right)^{1/2} \end{aligned}$$

and from Parseval's equality we have

$$\begin{aligned} \frac{1}{T} \sum_{s=0}^{T-1} \left| \sum_{j=0}^{t-1} \alpha_j^{(-\widehat{d})} e^{-i\frac{2\pi s}{T}j} \right|^2 &= \sum_{j=0}^{t-1} |\alpha_j^{(-\widehat{d})}|^2 \leq \sum_{j=0}^{\infty} |\alpha_j^{(-\widehat{d})}|^2 \\ &= \frac{\Gamma(1 - 2\widehat{d})}{(\Gamma(1 - \widehat{d}))^2} < \frac{\Gamma(\epsilon)}{(\Gamma(0.5(1 + \epsilon)))^2}. \end{aligned}$$

From the equality

$$\frac{1}{T} \sum_{s=0}^{T-1} |z_{\{\widehat{w}^* - \widehat{w}\}}^T(2\pi s/T)|^2 = \frac{1}{T} \sum_{t=1}^T \{\widehat{w}^*(t) - \widehat{w}(t)\}^2$$

it follows that

$$\frac{1}{T} \sum_{t=1}^T |\widehat{y}^*(t) - y(t)|^2 \leq \frac{\Gamma(\epsilon)}{(\Gamma(0.5(1 + \epsilon)))^2} \frac{1}{T} \sum_{t=1}^T \{\widehat{w}^*(t) - \widehat{w}(t)\}^2.$$

The foregoing relationships hold for all possible pairs $(y(t), \widehat{y}^*(t))$, $t = 1, \dots, T$, with probability one with respect to $(\Omega \otimes \Omega^*, \mathfrak{F} \otimes \mathfrak{F}^*, P(P^*))$ and we can therefore conclude that

$$E[E^*[\frac{1}{T} \sum_{t=1}^T |\widehat{y}^*(t) - y(t)|^2]] \leq GE[E^*[\frac{1}{T} \sum_{t=1}^T \{\widehat{w}^*(t) - \widehat{w}(t)\}^2]],$$

where $G = \Gamma(\epsilon)/(\Gamma(0.5(1 + \epsilon)))^2$. \blacksquare

Now let $\widehat{\mathbf{s}}_T^* = (\widehat{s}_{1T}^*, \dots, \widehat{s}_{mT}^*)'$, where $\widehat{s}_{iT}^* = s_i(\widehat{y}^*(1), \dots, \widehat{y}^*(T))$, $i = 1, \dots, m$, denote the value of the statistic of interest when calculated from a pre-filtered sieve bootstrap realization. Let $\widehat{\psi}_T^*(\boldsymbol{\tau}) = E^*[\exp(i\boldsymbol{\tau}'\widehat{\mathbf{s}}_T^*)]$ and set $\widehat{\boldsymbol{\zeta}}_T^* = \widehat{\mathbf{V}}_T^{*-1/2}T^{-1/2}(\widehat{\mathbf{s}}_T^* - E^*[\widehat{\mathbf{s}}_T^*])$ where $\widehat{\mathbf{V}}_T^* = T^{-1}E^*[(\widehat{\mathbf{s}}_T^* - E^*[\widehat{\mathbf{s}}_T^*])(\widehat{\mathbf{s}}_T^* - E^*[\widehat{\mathbf{s}}_T^*])']$.

Lemma 10 *Suppose that the process $y(t)$ satisfies Assumptions 1 and 2, and that the statistic \mathbf{s}_T satisfies Assumption 4. Then for all \widehat{d} such that $\widehat{d} - d \in N_{\delta_T}$ where $\delta_T \rightarrow 0$ as $T \rightarrow \infty$, and $\widehat{d} \in (-0.5, 0.5(1 - \epsilon))$ where $\epsilon > 0$, $E[E^*[\|\widehat{\mathbf{s}}_T^* - \mathbf{s}_T\|^2]] = \exp(2\delta_T \log T)o(T^{-1+\beta})$ for all $\beta > 0$. Furthermore, if $\delta_T \log T \rightarrow 0$ as $T \rightarrow \infty$ then for all $\boldsymbol{\tau}$ such that $\|\boldsymbol{\tau}\| \leq T^{\beta/2}$, it follows that $|\widehat{\psi}_T^*(\boldsymbol{\tau}) - \psi_T(\boldsymbol{\tau})| = \exp(\delta_T \log T)o_p\{T^{-1/2+\beta}\}$ uniformly in $\boldsymbol{\tau}$.*

Proof: Since \mathbf{s}_T satisfies Assumption 4 there exists a constant $M < \infty$ such that $\|\widehat{\mathbf{s}}_T^* - \mathbf{s}_T\|^2 \leq mMT^{-1} \sum_{t=1}^T |(\widehat{y}^*(t) - y(t))|^2$, from which it immediately follows that $E[E^*[\|\widehat{\mathbf{s}}_T^* - \mathbf{s}_T\|^2]] \leq mM E[E^*[T^{-1} \sum_{t=1}^T |(\widehat{y}^*(t) - y(t))|^2]]$. But by Proposition 9 $E[E^*[T^{-1} \sum_{t=1}^T |(\widehat{y}^*(t) - y(t))|^2]] \leq GE[E^*[T^{-1} \sum_{t=1}^T |(\widehat{w}^*(t) - \widehat{w}(t))|^2]]$ where $G < \infty$, and a repetition of the argument used in the proof of Lemma 7 shows that $E[E^*[(\widehat{w}^*(t) - \widehat{w}(t))^2]] = o(T^{-(1-2\delta')+\beta})$ where $\delta' = \max\{0, d - \widehat{d}\} < \delta_T$ for all $\beta > 0$. We are therefore lead to the conclusion that $E[E^*[\|\widehat{\mathbf{s}}_T^* - \mathbf{s}_T\|^2]] = \exp(2\delta_T \log T)o(T^{-1+\beta})$. This proves the first part of the lemma. The proof of the second part of the lemma now follows that used in Lemma 7 in an obvious manner. \blacksquare

Theorem 11 *Suppose that the statistic \mathbf{s}_T satisfies Assumption 4 and Assumption 5 with $r \geq 4$ when calculated from any process $y(t)$ that satisfies Assumptions 1 and 2. Then for all \widehat{d} such that $\widehat{d} - d \in N_{\delta_T}$ where $\delta_T \log T \rightarrow 0$ as $T \rightarrow \infty$,*

$$\sup_{\mathbf{z}} |P^*(\widehat{\boldsymbol{\zeta}}_T^* \leq \mathbf{z}) - P(\boldsymbol{\zeta}_T \leq \mathbf{z})| = \exp(\delta_T \log T)O_p(T^{-1+\beta})$$

for all $\beta > 0$.

Proof: Apart from minor notational changes and an allowance for the filtering that occurs at Steps PFSBS1 and PFSBS3, the argument leading from Lemma 10 to Theorem 11 is almost identical to that leading from Lemma 7 to Theorem 6. The details are therefore omitted. \blacksquare

In practice, of course, the preliminary estimate \widehat{d} will be constructed from the data, and from Theorem 11 we can see that if $\widehat{d} - d \in N_{\delta_T}$ as $T \rightarrow \infty$, where $\delta_T \log T \rightarrow 0$, then the error of the pre-filtered sieve bootstrap will be $O_p(T^{-1+\beta})$ for all $\beta > 0$. Thus, if $|\widehat{d} - d| \log T \rightarrow 0$ a.s. as $T \rightarrow \infty$ then the pre-filtered sieve bootstrap will achieve a convergence rate arbitrarily close to the rate obtained with simple random samples.

To establish the required convergence result for \widehat{d} requires the establishment of both consistency and the appropriate (limiting) tail behaviour for the standardized estimator $N^{1/2}(\widehat{d} - d)$, where N is a monotonically increasing function of T such that $N/T \rightarrow 0$ as $T \rightarrow \infty$. In particular, if $N^{1/2}(\widehat{d} - d)$ were $\mathbb{N}(0, v)$ then it would follow from the tail area properties of the normal distribution that $\lim_{T \rightarrow \infty} P(|\widehat{d} - d| > \epsilon N^{-1/2+\delta}) \leq \exp(-\epsilon^2 N^{2\delta}/2v)$ for any δ , $0 < \delta < 0.5$ and $\epsilon > 0$. Since $\exp(-\epsilon N^\delta/2v) < |r|^{N^\delta}$ for all r such that $\exp(-\epsilon/2v) < |r| < 1$ we could then conclude from the Borel-Cantelli lemma that $N^{1/2-\delta}|\widehat{d} - d|$ converged to zero almost surely and hence that $|\widehat{d} - d| \log T = o(1)$ *a.s.* if $\log T/N^{1/2-\delta} \rightarrow 0$. Note that asymptotic Gaussianity (associated with a \sqrt{N} -CAN estimator of d) would not be sufficient here, as departures of $N^{1/2}(\widehat{d} - d)$ from zero that are negligible in the sense of weak convergence need not be so for large-deviation probabilities. Large-deviation type results can, of course, be formally established on a case by case basis. In particular, a corollary of [Giraitis and Robinson \(2003, Lemma 5.8\)](#) is that the semi-parametric local Whittle (SPWL) estimator satisfies $P(|\widehat{d} - d| \log T > \epsilon) = o(N^{-p})$, where $p > 1/\epsilon$ and N , the bandwidth, satisfies $T^\epsilon < N < T^{1-\epsilon}$ for some $\epsilon > 0$.

In the simulation exercise that follows we apply a PFSBS algorithm based on a pre-filtering value of d that is produced by bias correcting the SPWL estimator. The correction incorporates a combination of the analytical adjustment of [Andrews and Sun \(2004\)](#) and a sieve bootstrap-based bias adjustment, the latter justified on the basis of the Edgeworth expansion of [Giraitis and Robinson \(2003\)](#). In support of this choice of pre-filtering value we invoke the Monte Carlo evidence in [Poskitt, Martin and Grose \(2012\)](#) that demonstrates the accuracy of (different versions of) a bias-adjusted SPWL estimator, most notably in comparison with the raw SPWL estimator; see also the discussion in [Nielsen and Frederiksen \(2005\)](#). The simulation design adopted in the current paper is identical to that adopted in [Poskitt et al.](#) and the bias-adjusted SPWL estimator that minimized mean squared error across the Monte Carlo replications there, in any given design setting, is used here as the pre-filter.

5 Simulation Exercise

In this section we examine the performance of the sieve bootstrap techniques via a simulation experiment. Specifically, we investigate the accuracy with which both the raw and pre-filtered sieve algorithms approximate the sampling distributions of the sample mean, \bar{y}_T , and the k th-order sample autocorrelation coefficient, $\hat{\rho}(k)$, for $k = 1, 3, 6$ and 9 .

Regarding $\bar{y}_T = \sum_{t=1}^T y(t)/T$, various properties of this statistic are well known, and in the investigation of any bootstrap procedure an examination of its ability to mimic these is a natural focal point. In particular, theoretical (asymptotic) properties

notwithstanding, it is of interest to investigate the nature of the finite sampling performance of the sieve-based estimators of this important sampling distribution, and to document the extent of the improvement yielded by the pre-filtering process. The characteristics that are of particular interest in the context of fractionally integrated data are, first, that

$$\text{Var}[\bar{y}_T] = \frac{1}{T} \sum_{k=1}^{T-1} \left(1 - \frac{|k|}{T}\right) \gamma(k), \quad (5.1)$$

second, that $\text{Var}[\bar{y}_T] \sim T^{2d-1} \omega^2$ where

$$\omega^2 = \frac{\{\sigma \kappa(1)\}^2 \Gamma(1-2d)}{(1+2d)\Gamma(1+d)\Gamma(1-d)} \quad (5.2)$$

as $T \rightarrow \infty$, and third, that the re-normalized mean $T^{1/2-d}(\bar{y}_T - \mu) \xrightarrow{D} \mathbb{N}(0, \omega^2)$, where $E[y_t] = \mu$; see [Hosking \(1996, Theorem 8\)](#). In the case where the simulated data is Gaussian all semi-invariants of \bar{y}_T of order greater than two are zero, of course, and the terms in the Edgeworth expansion in [\(3.1\)](#) beyond the first are null. Given knowledge of the true sampling variance of the mean in [\(5.1\)](#), the representativeness of the Monte Carlo (MC) distribution, the relevance of the asymptotic approximation and the accuracy of the bootstrap methods can all be assessed against the exact Gaussian sampling distribution.²

With regard to the k^{th} sample autocorrelation coefficient, defined here as³

$$\hat{\rho}(k) = \frac{\sum_{t=1}^{T-k} (y(t) - \bar{y}_T)(y(t+k) - \bar{y}_T)}{\sum_{t=1}^T (y(t) - \bar{y}_T)^2}, \quad (5.3)$$

the finite sample distribution under long memory is unknown. However, the relevant asymptotic results are documented in [Hosking \(1996\)](#), with asymptotic normality shown to hold for the appropriately standardized statistic for $d \leq 0.25$, and Hosking's "modified" Rosenblatt distribution being the relevant limiting distribution for $0.25 < d < 0.5$. Hence, it is of interest to explore the performance of the sieve-based techniques in replicating (Monte Carlo estimates of) the finite sampling distributions in the two regions of the parameter space in which the asymptotic behaviour of $\hat{\rho}(k)$ differs.

Further, [Lieberman et al. \(2001\)](#) ("LRZ" hereafter) provide the analytical details

² [Andrews et al. \(2006\)](#) remark that in the Gaussian case "the sample mean is an unbiased estimator of μ with an exact normal distribution, which can be used to develop inference concerning μ ", but they make no mention of issues associated with estimating the sampling variance of \bar{y}_T . We should perhaps point out that the sample mean is not the best linear unbiased estimator of μ for a fractional process, see [Adenstedt \(1974\)](#). [Adenstedt \(1974, Theorem 5.2\)](#) presents an alternative estimator that is asymptotically efficient, albeit infeasible in practice because it is a function of the unknown d . We thank the Editor for bringing this paper to our attention.

³ There are several closely-related ways to define the sample autocorrelation; we have used Hosking's (1996) specification in our work here.

of the Edgeworth expansion for the distribution of $\sqrt{T}(\hat{\rho}_0(k) - \rho(k))$, where

$$\hat{\rho}_0(k) = \frac{\sum_{t=1}^{T-k} y(t)y(t+k)}{\sum_{t=1}^T y(t)^2} \quad (5.4)$$

and the mean μ is known to be zero. As well as being derived for a statistic that both assumes and imposes a known zero mean, the LRZ expansion is dependent on the true (unknown) values of the ARFIMA parameters, as well as being valid only for very small values of d ; all such features limiting its empirical applicability. However, in the current experimental setting it serves as a very useful check of the performance of the bootstrap method. A good match in the case where the expansion is applicable suggests that the bootstrap method(s) may also perform well in the parts of the parameter space in which the Edgeworth approach is inapplicable, and in empirical scenarios in which parameters are (of course) unknown. For convenience, and to facilitate the reproducibility of our results, we document in the Appendix the details of the Edgeworth formula applied here.

5.1 Simulation Design

Data are simulated from a zero-mean Gaussian ARFIMA(p, d, q) process, with autoregressive lag order $p = 1$ and moving average lag order $q = 0$,

$$(1 - L)^d(1 - \phi L)y(t) = \varepsilon(t), \quad (5.5)$$

where $1 - \phi z$ is the operator for a stationary AR(1) component and $\varepsilon(t)$ is zero-mean Gaussian white noise. The theoretical autocovariance function (ACF) for this process can be computed using the procedures of [Sowell \(1992\)](#). The process in (5.5) is simulated $R = 1000$ times for $d = 0.0, 0.2, 0.3, 0.4$; $\phi = 0.3$ and 0.6 ; and for various sample sizes T , via the Levinson recursion applied to the ACF of the desired ARFIMA(1, d , 0) process and the generated pseudo-random $\varepsilon(t)$ (see, for instance, [Brockwell and Davis, 1991](#), §5.2). The ARFIMA ACF for given T , ϕ and d is calculated using Sowell's (1992) algorithm as modified by [Doornik and Ooms \(2001\)](#).

For each realization r of the process we compute the relevant statistic, $s_{T,r}$, plus $B = 1000$ estimates $s_{T,r(b)}^*$, constructed using $b = 1, \dots, B$ bootstrap re-samples obtained via the relevant bootstrap algorithm. Each realized value $s_{T,r}$ thus has associated with it a ‘‘bootstrap distribution’’ based on the B bootstrap resamples $s_{T,r(b)}^*$, $b = 1, \dots, B$, with each such distribution serving as an estimate of the sampling distribution of s_T . In order to assess the R bootstrap distributions against a comparator distribution of s_T – whether that be the known finite sample distribution (as in the case of \bar{y}_T), the finite sample distribution estimated from the Monte Carlo draws, or an Edgeworth approximation – we first compute an ‘‘average’’ bootstrap distribution

by sorting the B bootstrap draws for each MC replication into ascending order, then averaging these ordered bootstrap values across the Monte Carlo draws. The B averaged draws are then used to produce a kernel density estimate, which we refer to as the average bootstrap distribution. This bootstrap estimate, when plotted against (a representation of) the true distribution of s_T , allows for a direct visual assessment of the overall accuracy of the bootstrap distributions⁴.

Following common practice (Politis, 2003, §3), the order of the autoregressive approximation in the sieve was set to $h = \hat{h}_T = \operatorname{argmin}_{h=0,1,\dots,M_T} \log(\hat{\sigma}_h^2) + 2h/T$, where $\hat{\sigma}_h^2$ denotes the residual mean square obtained from an $AR(h)$ model and $M_T = \lceil (\log T)^2 \rceil$. Let $\bar{h}_T = \operatorname{argmin}_{h=0,1,\dots,M_T} L_T(h)$ where $L_T(h) = (\sigma_h^2 - \sigma^2) + h\sigma^2/T$ and σ^2 and σ_h^2 are as defined in (1) and (2.4) respectively. The function $L_T(h)$ was introduced by Shibata (1980) as a figure of merit and the $AR(\hat{h}_T)$ model is asymptotically efficient in the sense that $L_T(\hat{h}_T) = L_T(\bar{h}_T)\{1 + o(1)\}$ *a.s.* as $T \rightarrow \infty$ (Poskitt, 2007, Theorem 9). It follows that $\hat{h}_T/\bar{h}_T \rightarrow 1$ *a.s.* as $T \rightarrow \infty$, so as T increases \hat{h}_T behaves almost surely like a deterministic sequence that satisfies the previous technical requirements. We can therefore conclude that, although the use of *AIC* introduces an added element of randomness, the results of Section 3 and Section 4 – in which h is treated as fixed – will still hold true.

In the case of \bar{y}_T we supplement the graphical results by tabulating the ratio of the average bootstrap estimate of the sampling variance of \bar{y}_T (for both bootstrap methods) to its true sampling variance as per (5.1). For the sample autocorrelations – for which the finite sample distribution is completely unknown – we focus on the accuracy with which the bootstrap methods reproduce the distribution as a whole; measuring the “closeness” of the averaged bootstrap distribution (in any particular case) to the chosen comparator distribution using three goodness of fit measures. For example, denoting the ordinates of the Monte Carlo comparator and the (averaged) bootstrap-based probability density functions (pdfs) at the j^{th} (sorted) Monte Carlo realization value as $p_{mc}(s_j)$ and $p_{bs}(s_j)$ respectively, we calculate

$$RMSD = \sqrt{\frac{1}{R} \sum_{j=1}^R (p_{mc}(s_j) - p_{bs}(s_j))^2}$$

⁴ To document the extent of the variation in the bootstrap samples (and hence density estimates) across Monte Carlo draws, we also produced kernel density estimates based on the quartiles of the ordered bootstrap iterates. That is, if $s_{T,r}^*(b_j)$, $r = 1, \dots, R$, is the set of R j^{th} -largest bootstrap values, for which we calculate the q^{th} quantile, then the collection of B such quantiles is an estimate of the “distribution” to the left of which proportion q of the R BS distributions lie. Inspection of these bootstrap quantiles did not indicate a great deal of variation in the bootstrap distributions across Monte Carlo draws.

and the Kullback-Leibler divergence

$$KLD = \sum_{j=1}^R p_{mc}(s_j) \ln \frac{p_{mc}(s_j)}{p_{bs}(s_j)} \simeq \frac{1}{R} \sum_{j=1}^R \ln \frac{p_{mc}(s_j)}{p_{bs}(s_j)},$$

where this last follows given s_j is by definition a random draw from the Monte Carlo distribution of s_T . We also produce a ‘‘GINI’’-style statistic by estimating the cumulative distribution functions, $P_{mc}(s_j)$ and $P_{bs}(s_j)$, $j = 1, 2, \dots, J$, in the same manner as the pdfs above, then calculating the area between the resulting PP-plot and the line of equality via a numerical (trapezoidal) estimate of

$$A = \int_0^1 |P_{bs}(s) - P_{mc}(s)| dP_{mc}(s).$$

The GINI coefficient is then just $2A$.

5.2 Simulation results: sample mean

Figure 1 graphs the distribution of $T^{1/2-d}(\bar{y}_T - \mu)$ (where $\mu = 0$) observed across the Monte Carlo draws (denoted by MC), the averaged (raw) sieve bootstrap distribution of $T^{1/2-d}(\bar{y}_T^* - \bar{y}_T)$ (denoted by SBS), and the exact $\mathbb{N}(0, \bar{\omega}^2)$ distribution, with $\bar{\omega}^2 = T^{1-2d}Var[\bar{y}_T]$, for $T = 500$, $\phi = 0.6$, and $d = 0, 0.2, 0.3, 0.4$. We have suppressed the plot of the asymptotic $\mathbb{N}(0, \omega^2)$ distribution of $T^{1/2-d}(\bar{y}_T - \mu)$ since at this sample size it is virtually indistinguishable from the exact.

When $d = 0$ we can see that all three distributions are very nearly identical. When $d > 0$, however, it is clear that the variance of \bar{y}_T is substantially underestimated by the bootstrap procedure. This result is further confirmed by inspection of Table 1, which reports the ratio of the average SBS estimate of the standard deviation of \bar{y}_T to the exact standard deviation given by the square root of (5.1), for $T = 100$ and 500 , and for $\phi = 0.3$ and 0.6 . The underestimation for $d > 0$ is very marked for both values of ϕ and both sample sizes with, indeed, the degree of underestimation increasing with ϕ and there being no uniform tendency for improvement as the sample size increases⁵.

The reason for the underestimation stems from the fact that the raw sieve bootstrap variance is

$$Var^*[\bar{y}_T^*] = \frac{1}{T} \sum_{k=1-T}^{T-1} \left(1 - \frac{|k|}{T}\right) \bar{\gamma}_h(k),$$

where $\bar{\gamma}_h(k) = \hat{\gamma}(k)$, $k = 0, 1, \dots, h$, and $\sum_{j=0}^h \bar{\phi}_h(j) \bar{\gamma}_h(k-j) = 0$, $k = h+1, \dots$, with

⁵ The mean and skewness of the re-normalised difference $T^{1/2-d}(\bar{y}_T^* - \bar{y}_T)$ are close to zero, and the kurtosis is approximately 3. Thus it is only the underestimation of variance that presents a problem. A similar phenomenon with the block bootstrap was observed previously by Hesterberg (1997). Hesterberg offers no explanation for its occurrence, but simply suggests that estimating the variance of the sample mean is substantially more difficult in the long memory case than it is for short memory processes.

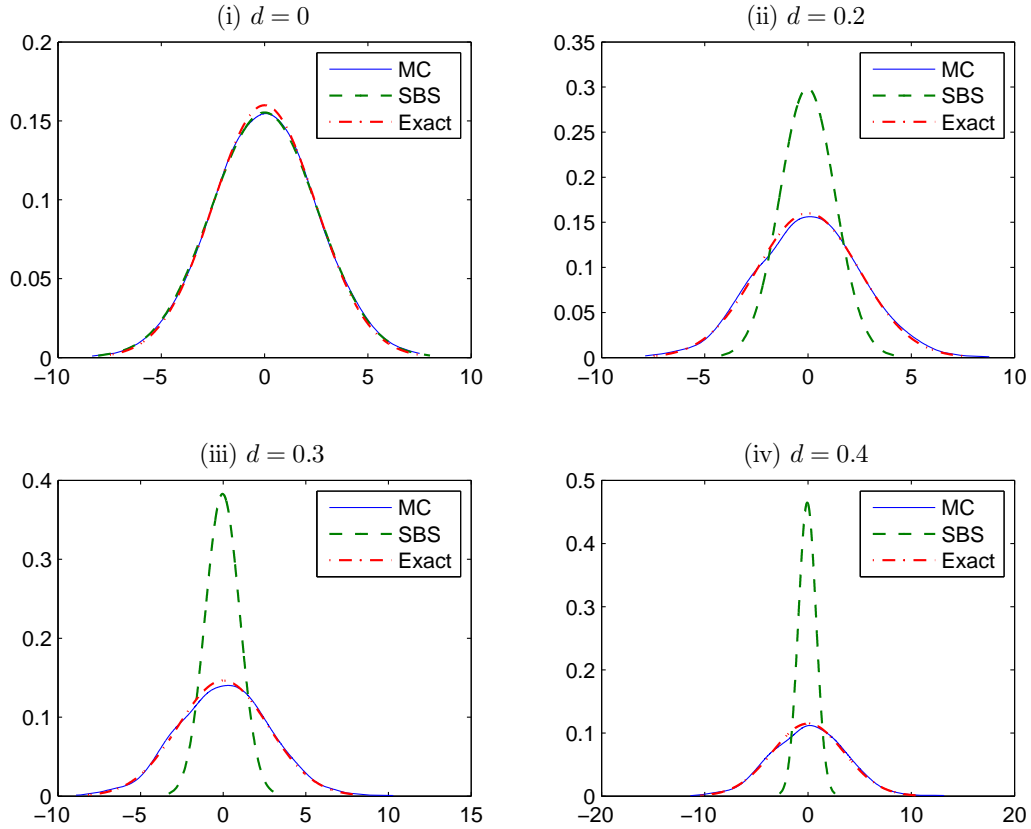


Figure 1: Densities of the re-normalized sample mean under ARFIMA(1, d , 0) with $T = 500$, $\phi = 0.6$, and $d = 0, 0.2, 0.3, 0.4$: Monte Carlo (MC); averaged (raw) sieve bootstrap (SBS) and exact ($N(0, \bar{\omega}^2)$), where $\bar{\omega}^2 = T^{1-2d} \text{Var}[\bar{y}_T]$, with $\text{Var}[\bar{y}_T]$ as given in (5.1).

$\hat{\gamma}(k) = \frac{1}{T-k} \sum_{t=1}^{T-k} (y(t) - \bar{y}_T)(y(t+k) - \bar{y}_T)$; and Hosking (1996) shows that the $\hat{\gamma}(k)$ can have substantial negative bias relative to the corresponding true values even for moderate to large samples, particularly when d is large.

This phenomenon is illustrated in Figure 2, which depicts the theoretical autocovariance function and the value of $\hat{\gamma}(k)$ for $k = 0, \dots, 100$ obtained from samples of size $T = 1000$, computed from two fractional noise (ARFIMA(0, d , 0)) processes with $d = 0.3$ and $d = 0.4$, and averaged across the R replications. Hosking (1996, Theorem 3) provides the following formula for the asymptotic bias of the autocovariances

$$E[\hat{\gamma}(k) - \gamma(k)] \sim -\omega^2 T^{2d-1}, \quad (5.6)$$

(with ω^2 as defined in 5.2), which depends on d but is independent of k . This feature is reflected in the simulated sample bias, computed as the difference between the mean of the simulated sample autocovariances at each lag k and the corresponding true $\gamma(k)$. For $T = 1000$ this estimated bias is in close accord with the asymptotic approximation

TABLE 1

Standard deviation of \bar{y}_T : averaged SBS estimate as a percentage of $\sqrt{\text{Var}[\bar{y}_T]}$, with $\text{Var}[\bar{y}_T]$ as defined in (5.1).

		d			
		0.0	0.2	0.3	0.4
ϕ	T				
0.3	100	95.6%	57.2%	42.6%	28%
	500	99.2%	48.6%	35.1%	22.8%
0.6	100	93.3%	60.2%	46%	31%
	500	99.1%	51.5%	36.8%	23.8%

in (5.6), as can be seen in Figure 2⁶.

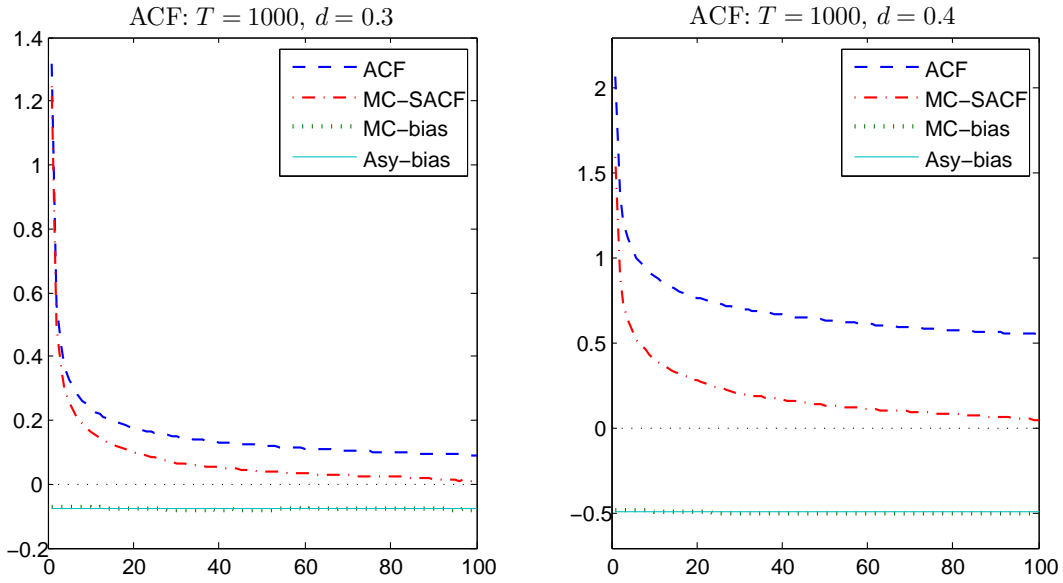


Figure 2: Theoretical autocovariance function (ACF); Monte-Carlo estimate of the expected sample autocovariance function (MC-SACF); Monte-Carlo estimate of the bias in the SACF (MC-Bias); and the asymptotic bias as per (5.6) (Asy-Bias), under fractional noise with $d = 0.3, 0.4$ and $T = 1000$.

It is of interest then to ascertain whether the feasible PFSBS algorithm, in *implicitly* producing more accurate estimates of the $\gamma(k)$ in the process of yielding bootstrap draws of \bar{y}_T , (via the application of the sieve to a shorter memory process) yields

⁶ Noting that $\frac{1}{T} \sum_{k=1}^{T-1} (1 - \frac{|k|}{T}) \equiv 1$, it is apparent from equation (5.6) that the addition of $\omega^2 T^{2d-1}$ to the bootstrap variance would provide an asymptotically valid (albeit empirically infeasible) correction that would compensate for the bias of the $\hat{\gamma}(k)$ and the underestimation of the true persistence in the observed process.

an estimated sampling distribution for the mean with a variance that is closer to the theoretical value. As noted in Section 4, the PFSBS algorithm is based on a pre-filtering value of d that is deemed to be “optimal” in the matching experimental design in Poskitt et al. (2012).

Figure 3 graphs the Monte Carlo distribution of $T^{1/2-d}(\bar{y}_T - \mu)$, the PFSBS distribution of $T^{1/2-d}(\bar{y}_T^* - \bar{y}_T)$, and the $N(0, \bar{\omega}^2)$ distribution, for $T = 500$, $\phi = 0.6$, and $d = 0, 0.2, 0.3, 0.4$. We see that, despite a tendency to over-estimate $Var(\bar{y}_T)$ for

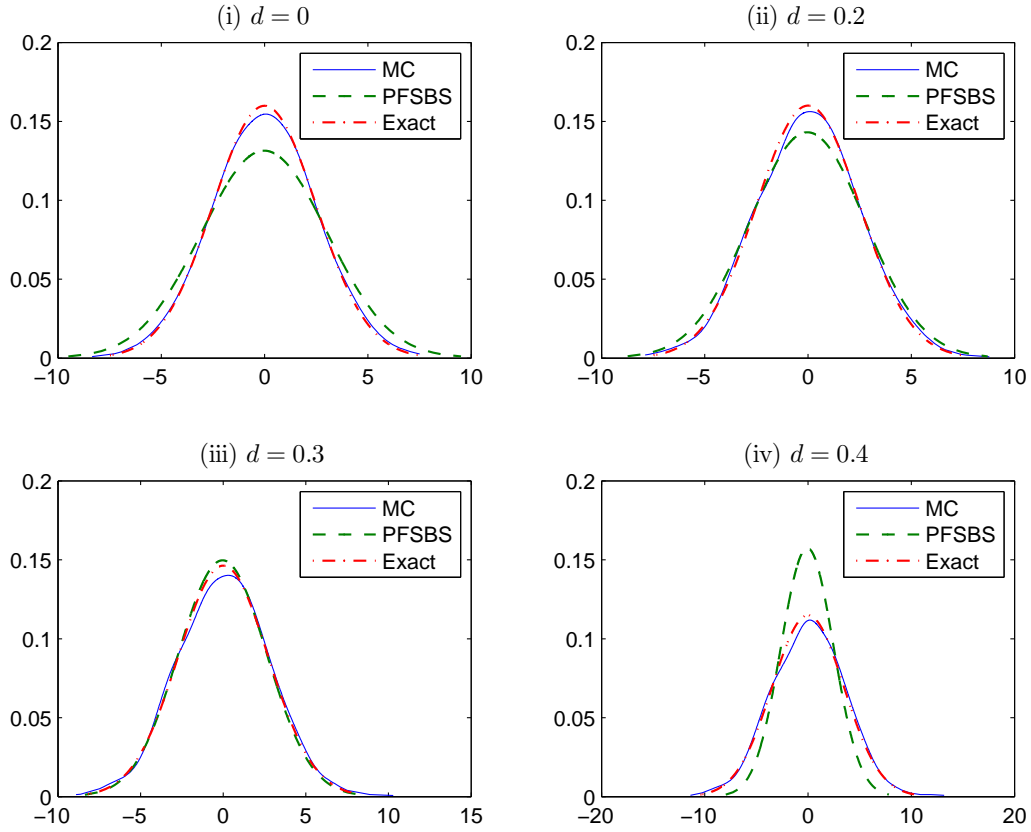


Figure 3: Densities of the re-normalized sample mean under ARFIMA(1, d , 0) with $T = 500$, $\phi = 0.6$, and $d = 0, 0.2, 0.3, 0.4$: Monte Carlo (MC); averaged pre-filtered sieve bootstrap (PFSBS) and exact ($N(0, \bar{\omega}^2)$), where $\bar{\omega}^2 = T^{1-2d}Var[\bar{y}_T]$, with $Var[\bar{y}_T]$ as given in (5.1).

$d = 0$ and under-estimate for large d ($d = 0.4$), the PFSBS results are far superior to those associated with the raw SBS, and reasonably close overall to the true sampling distribution. The averaged PFSBS estimate of $\sqrt{Var(\bar{y}_T)}$ as a percentage of the true $\sqrt{Var(\bar{y}_T)}$ is presented in Panel A of Table 2, for the two values of ϕ , $\phi = 0.3$ and 0.6 , and for $T = 100$ and 500 . The reasonable accuracy observed visually in Figure 3 for $\phi = 0.6$, for the larger sample size in particular, is broadly replicated for $\phi = 0.3$, $T = 500$, auguring well for the automated use of the pre-filtering method in practice.

As a final point here, it is of interest to ascertain the performance of the PFSBS technique in which we simply *assign* a value to d with which to pre-filter, rather

TABLE 2

Standard deviation of \bar{y}_T : averaged PFSBS and FPSBS estimates as a percentage of $\sqrt{\text{Var}[\bar{y}_T]}$, with $\text{Var}[\bar{y}_T]$ as defined in (5.1).

Panel A: PFSBS					
		d			
		0.0	0.2	0.3	0.4
ϕ	T				
0.3	100	141.2%	125.3%	109.8%	84.3%
	500	116.6%	106.9%	93.4%	69.9%
0.6	100	158.7%	142.9%	127.0%	100.4%
	500	117.1%	107.5%	94.0%	70.1%

Panel B: FPFBS					
		d			
		0.0	0.2	0.3	0.4
ϕ	T				
0.3	100	349.0%	191.1%	127.8%	74.3%
	500	573.4%	274.9%	171.6%	88.6%
0.6	100	316.0%	169.5%	115.7%	69.7%
	500	582.9%	268.6%	162.5%	84.7%

than selecting a particular estimator for this role.⁷ A fairly obvious choice is to set the pre-filtering value (d^f say) at 0.5; as the true d (in the experimental setting) is never greater than this it follows that imposing $d^f = 0.5$ results in a filtered series for which the effective fractional integration is always negative, and the filtered process of intermediate memory as a consequence. The estimates of the sieve parameters will therefore converge at the best possible rate $O(hT^{-1} \log T)$ as per Theorem 3, although d^f will obviously *not* satisfy the convergence properties outlined in the discussion following Theorem 11. We refer to this approach below as the “fixed pre-filtered bootstrap” (FPFBS).

As we see from Figure 4, the FPFBS, unsurprisingly, works reasonably well when d is large; i.e., for $d = 0.4$. For the smaller values for d , on the other hand, it works very poorly, resulting in an averaged bootstrap distribution for \bar{y}_T that is a very inaccurate match for the true distribution. In particular, as seen in panels (i) – (iii) of the Figure, and in Panel B of Table 2, the dispersion of the FPFBS-based distribution is much

⁷ The idea of imposing a “fixed” pre-filter arose out of a referee’s comment on an earlier version of the paper.

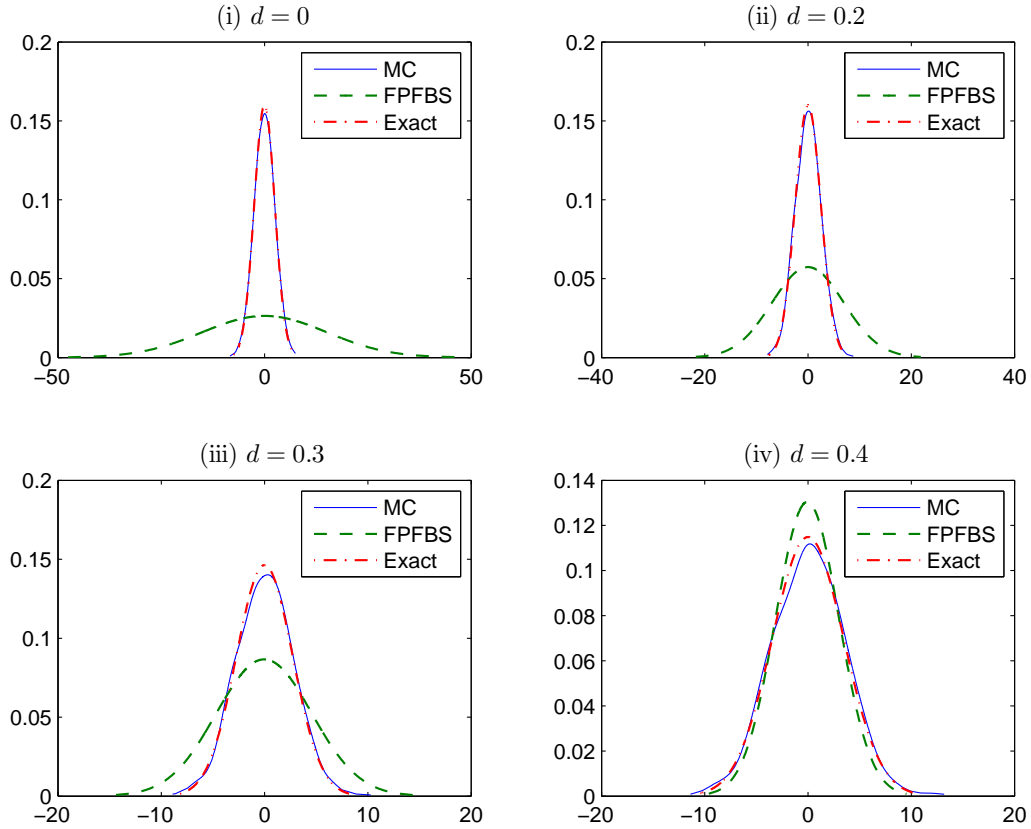


Figure 4: Densities of the re-normalized sample mean under ARFIMA(1, d , 0) with $T = 500$, $\phi = 0.6$, and $d = 0, 0.2, 0.3, 0.4$: Monte Carlo (MC); averaged fixed pre-filtered sieve bootstrap (FPFBS) and exact ($N(0, \bar{\omega}^2)$), where $\bar{\omega}^2 = T^{1-2d} \text{Var}[\bar{y}_T]$, with $\text{Var}[\bar{y}_T]$ as given in (5.1).

larger than that of the exact distribution, with the discrepancy increasing with the distance $|d - d^f|$. In short, it appears that fixing the pre-filter is not useful as a default setting, at least as regards estimating the distribution of \bar{y}_T .

5.3 Simulation results: sample autocorrelation

We begin by plotting various estimates of the true finite sampling distribution of $\hat{\rho}_0(k)$, for $k = 1, 3, 6$ and 9 , where the subscript 0 is used to emphasize that a mean of zero (for y_t) is both assumed and imposed in the calculation of the statistic (see (5.4)). We consider this particular version of the autocorrelation coefficient(s) in this initial exercise so as to enable the LRZ expansion (derived for this version) to be used as a comparator. The expansion is valid for $d < 0.1$ only (see Appendix); hence we conduct the comparison for a value of d in this range: $d = 0.08$. Results for $\phi = 0.3$ and $\phi = 0.6$ are presented in Figures 5 and 6 respectively, with $T = 500$ in both cases.

As is evident from inspection of the two graphs, the (raw) SBS estimate of the

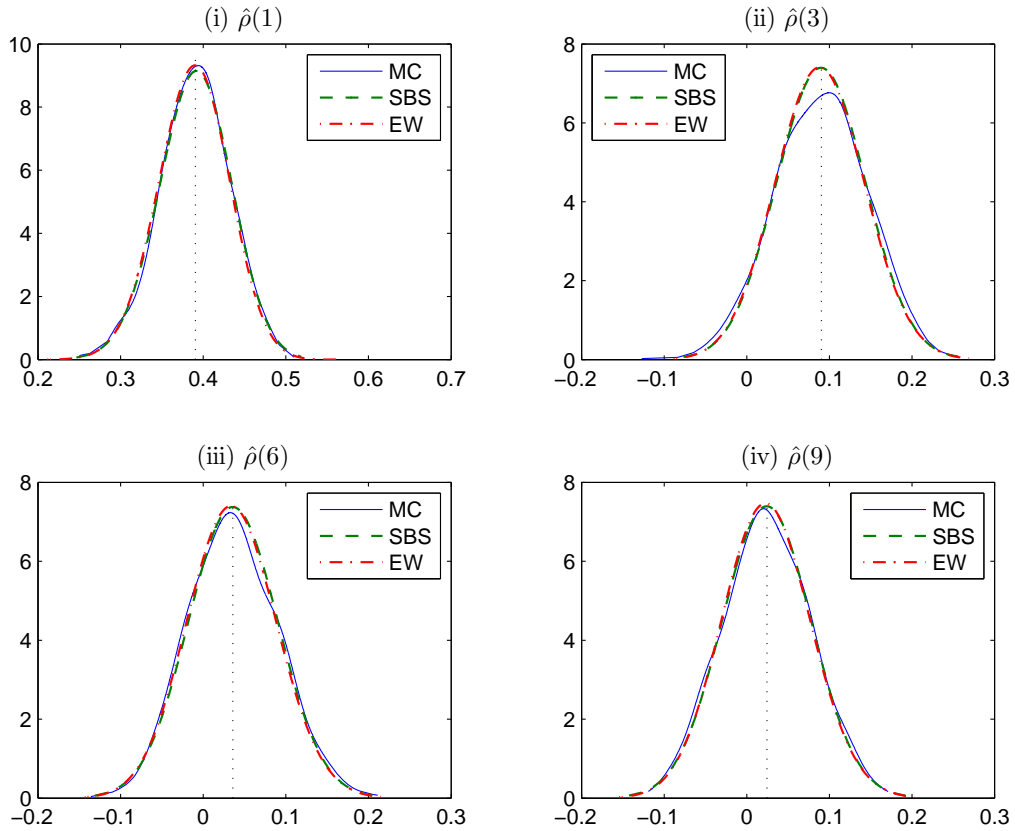


Figure 5: Densities of the zero-mean sample autocorrelation coefficient $\hat{\rho}_0(k)$, $k = 1, 3, 6, 9$: under ARFIMA(1, d , 0) with $T = 500$, $d = 0.08$, $\phi = 0.3$: Monte Carlo (MC); averaged (raw) sieve bootstrap (SBS); Edgeworth approximation (EW). The vertical dotted line indicates the position of the true value of $\rho(k)$, $k = 1, 3, 6, 9$.

distribution of $\hat{\rho}_0(k)$ is visually indistinguishable from the Edgeworth distribution⁸, with both being very similar to the Monte Carlo based estimate.⁹ As such, we conclude that when a finite sample comparator is available (i.e. under the conditions required for that comparator to be valid) the sieve bootstrap method is remarkably accurate. This gives one confidence in the ability of the bootstrap to provide an accurate result in the usual case in which such a comparator is unavailable.

In Figures 7–10 we proceed to document the performance of the two bootstrap methods, SBS and PFSBS, in regions of the parameter space in which the Edgeworth expansion is not valid, and the only comparator is the Monte Carlo-based estimate of the exact sampling distribution. We also include plots of the average FPFBS distrib-

⁸ The Edgeworth (EW) distribution plotted here has been re-centered on the true $\rho(k)$, and rescaled to remove the \sqrt{T} normalization of the expansion. See the Appendix for details.

⁹ We have reproduced results here based on 1000 replications in order to have all results comparable throughout the paper. In particular, due to the computational burden associated with the PFSBS methods, 1000 was a manageable choice for a general replication number. However, the results documented in Figures 5 and 6 have also been run using 10,000 replications, at which point the Monte Carlo estimate of the pdf is visually indistinguishable from the other two estimates.

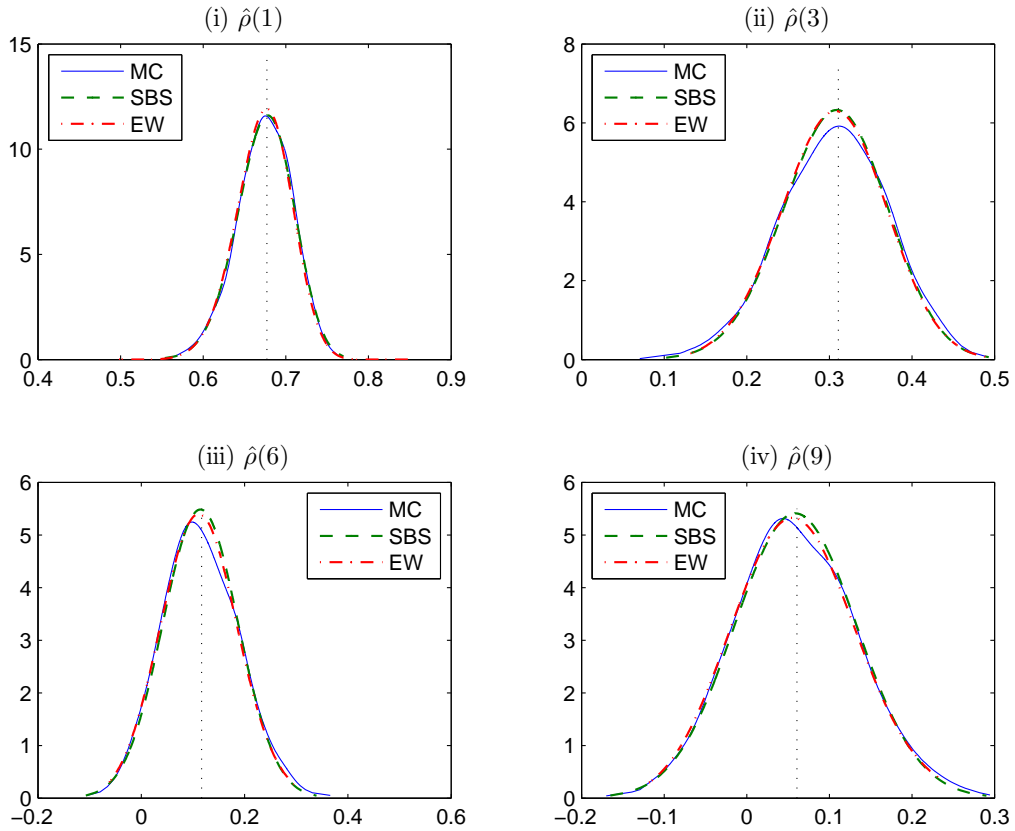


Figure 6: Densities of the zero-mean sample autocorrelation coefficient $\hat{\rho}_0(k)$, $k = 1, 3, 6, 9$: under ARFIMA(1, d , 0) with $T = 500$, $d = 0.08$, $\phi = 0.6$: Monte Carlo (MC); averaged (raw) sieve bootstrap (SBS); Edgeworth approximation (EW). The vertical dotted line indicates the position of the true value of $\rho(k)$, $k = 1, 3, 6, 9$.

utions calculated using, as in the previous section, $d^f = 0.5$. The distribution for the sample autocorrelation coefficient in (5.3) is now the one documented, for $k = 1, 3, 6$ and 9 , and the two scenarios considered are that in which asymptotic normality holds ($d \leq 0.25$), and that in which it does not ($d > 0.25$), with the modified Rosenblatt distribution being the relevant limiting form in the latter case. Specifically, in Figures 7 and 8, $d = 0.2$ and $\phi = 0.3$ and $\phi = 0.6$ respectively, whilst in Figures 9 and 10, $d = 0.4$ and $\phi = 0.3$ and $\phi = 0.6$ respectively. In order to supplement these graphical results, the measures of fit (as described in Section 5.1) are recorded in both panels of Table 3, in relative terms. That is, in Panel A, each fit measure for PFSBS is presented as a ratio to the corresponding fit measure for the raw sieve method (SBS); hence, a value less than one indicates that the pre-filtering yields a distribution that is a better fit to the Monte Carlo-based distribution. The corresponding results for FPFBS are presented in Panel B. Results recorded in all four figures and both panels of the table are for $T = 500$.

A visual inspection of the graphs in Figures 7 and 8 suggests that when the long

memory parameter is small ($d = 0.2$), and for both values of ϕ , the two bootstrap methods, SBS and PFSBS, provide reasonable accuracy. There is no clear cut supe-

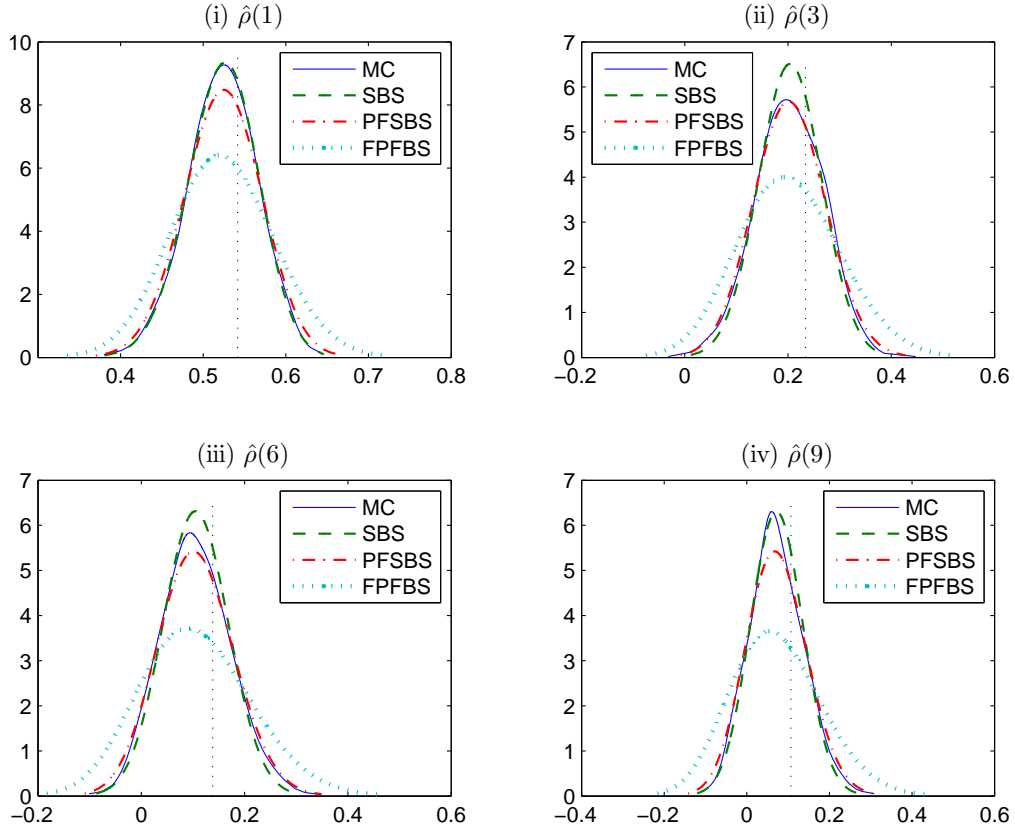


Figure 7: Densities of the sample autocorrelation coefficient $\hat{\rho}(k)$, $k = 1, 3, 6, 9$: under ARFIMA(1, d , 0) with $T = 500$, $d = 0.2$, $\phi = 0.3$: Monte Carlo (MC); averaged (raw) sieve bootstrap (SBS); averaged pre-filtered sieve bootstrap (PFSBS); averaged fixed pre-filtered sieve bootstrap (FPFBS). The vertical dotted line indicates the position of the true value of $\rho(k)$, $k = 1, 3, 6, 9$.

rriority of one method over the other, with the raw sieve method being superior to the PFSBS method for $k = 1$ and 9, and the opposite result obtaining for $k = 3$ and 6. These visual results on relative performance are confirmed (overall) by the numerical results in Table 3, Panel A, with virtually all ratios (associated with all three measures of fit) being greater than one (indicating the superiority of the raw method) for $k = 1$ and 9, and less than one for $k = 3$ and 6.

In contrast, for $d = 0.4$, the performances of the two methods are more distinct, with all graphs reproduced in Figures 9 and 10 - allied with the numerical results reported in Panel A, Table 3 - confirming the marked superiority of the PFSBS method in this part of the parameter space. Taken together these two sets of results suggest that a conservative approach to estimating the sampling distribution of the autocorrelation coefficients in empirical settings is to undertake the pre-filtering; the increase in accuracy in the long memory region being worth the slight reduction that *may* occur

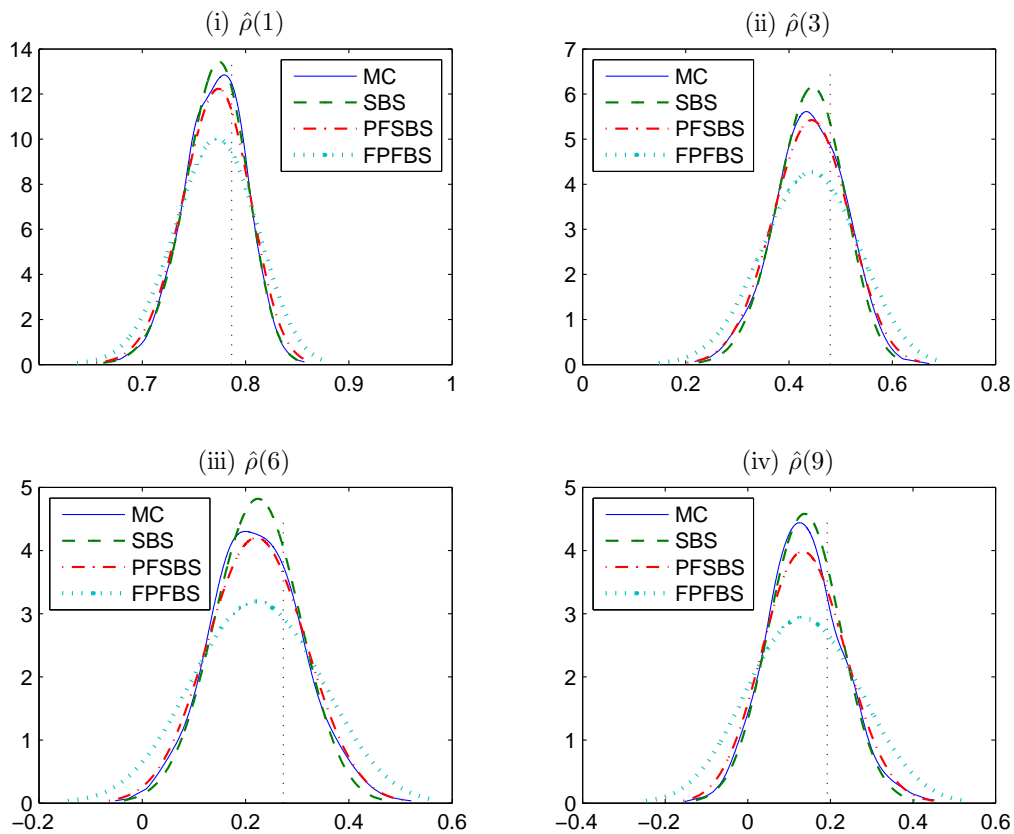


Figure 8: Densities of the sample autocorrelation coefficient $\hat{\rho}(k)$, $k = 1, 3, 6, 9$: under ARFIMA(1, d , 0) with $T = 500$, $d = 0.2$, $\phi = 0.6$: Monte Carlo (MC); averaged (raw) sieve bootstrap (SBS); averaged pre-filtered sieve bootstrap (PFSBS); averaged fixed pre-filtered sieve bootstrap (FPFBS). The vertical dotted line indicates the position of the true value of $\rho(k)$, $k = 1, 3, 6, 9$.

(relative to the raw sieve) if the true value of d is small.

Turning, finally, to the FPFBS, from inspection of Figures 7-10 and the results recorded in Panel B of Table 3, we see that while it virtually always outperforms the raw SBS for the larger value of d ($d = 0.4$), it does very poorly for $d = 0.2$. In the latter case we observe a “divergence” (as measured by our three goodness of fit measures) from the Monte Carlo distribution we are attempting to replicate that is several times larger than that of SBS; more than 100 times larger in one case. Most importantly, if we compare the FPFBS *directly* to the PFSBS (by making the appropriate simple calculations using the numbers recorded in the two panels of Table 3) we see that the FPFBS *never* outperforms the PFSBS, with the goodness of fit measures for the former ranging from (approximately) twice to seventeen times those of the latter. This poor performance (overall) of the FPFBS mimics that documented for the sample mean.

TABLE 3

Goodness of fit measures for $\hat{\rho}(k)$, $k = 1, 3, 6, 9$:
 measures for PFSBS and FPFBS relative to those for SBS.
 All results are for sample size $T = 500$.

		Panel A: PFSBS							
		$d = 0.2$				$d = 0.4$			
		Lag length k				Lag length k			
		1	3	6	9	1	3	6	9
RMSE	ϕ								
	0.3	4.566	0.384	0.616	1.276	0.235	0.173	0.358	0.374
	0.6	1.467	0.520	0.682	1.343	0.322	0.233	0.392	0.387
KLD	0.3	12.26	0.733	1.087	1.579	0.302	0.095	0.141	0.163
	0.6	6.357	1.304	0.858	1.310	0.500	0.130	0.195	0.231
GINI	0.3	9.006	0.339	0.605	1.128	0.264	0.128	0.234	0.333
	0.6	2.679	0.418	0.611	1.211	0.370	0.220	0.290	0.349
		Panel B: FPFBS							
		$d = 0.2$				$d = 0.4$			
		Lag length k				Lag length k			
		1	3	6	9	1	3	6	9
RMSE	ϕ								
	0.3	16.52	2.472	3.597	4.486	0.597	0.409	0.567	0.797
	0.6	4.385	2.416	2.699	4.257	0.597	0.436	0.530	0.662
KLD	0.3	102.99	10.32	18.79	18.18	0.971	0.380	0.426	0.507
	0.6	34.030	13.110	9.644	11.15	1.129	0.447	0.425	0.491
GINI	0.3	36.173	3.239	4.117	5.522	0.743	0.471	0.548	0.794
	0.6	9.979	2.946	3.220	4.829	0.759	0.502	0.519	0.655

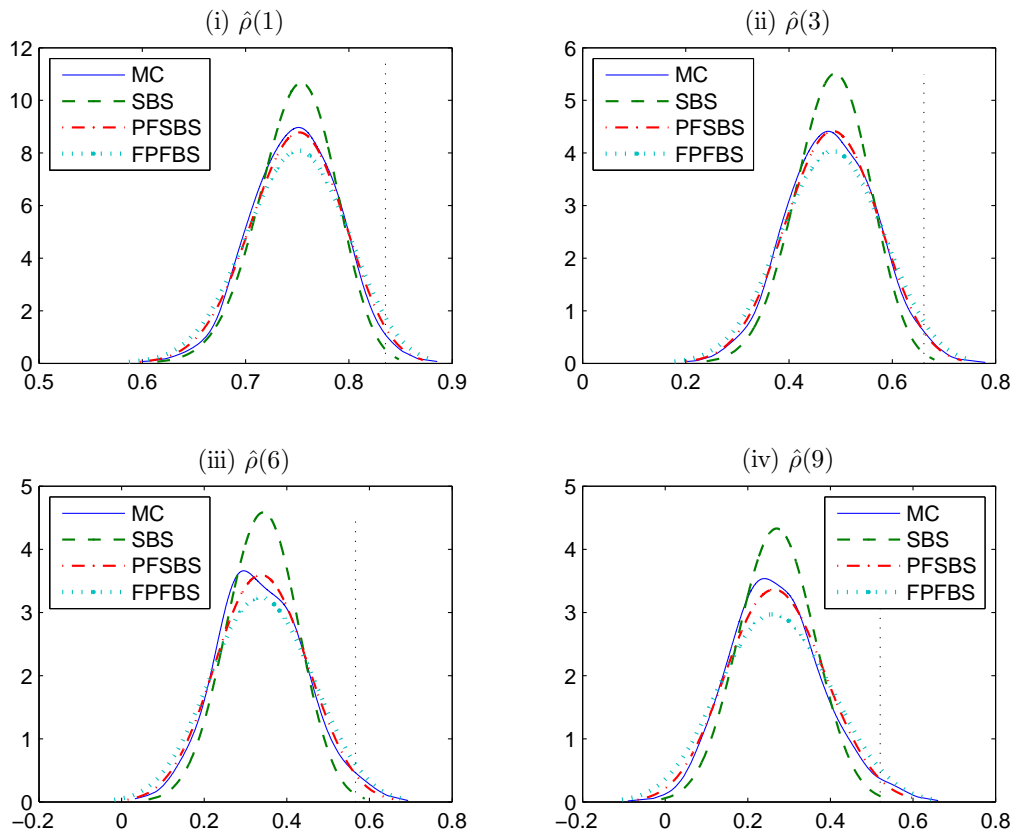


Figure 9: Densities of the sample autocorrelation coefficient $\hat{\rho}(k)$, $k = 1, 3, 6, 9$: under ARFIMA(1, d , 0) with $T = 500$, $d = 0.4$, $\phi = 0.3$: Monte Carlo (MC); averaged (raw) sieve bootstrap (SBS); averaged pre-filtered sieve bootstrap (PFSBS); averaged fixed pre-filtered sieve bootstrap (FPFBS). The vertical dotted line indicates the position of the true value of $\rho(k)$, $k = 1, 3, 6, 9$.

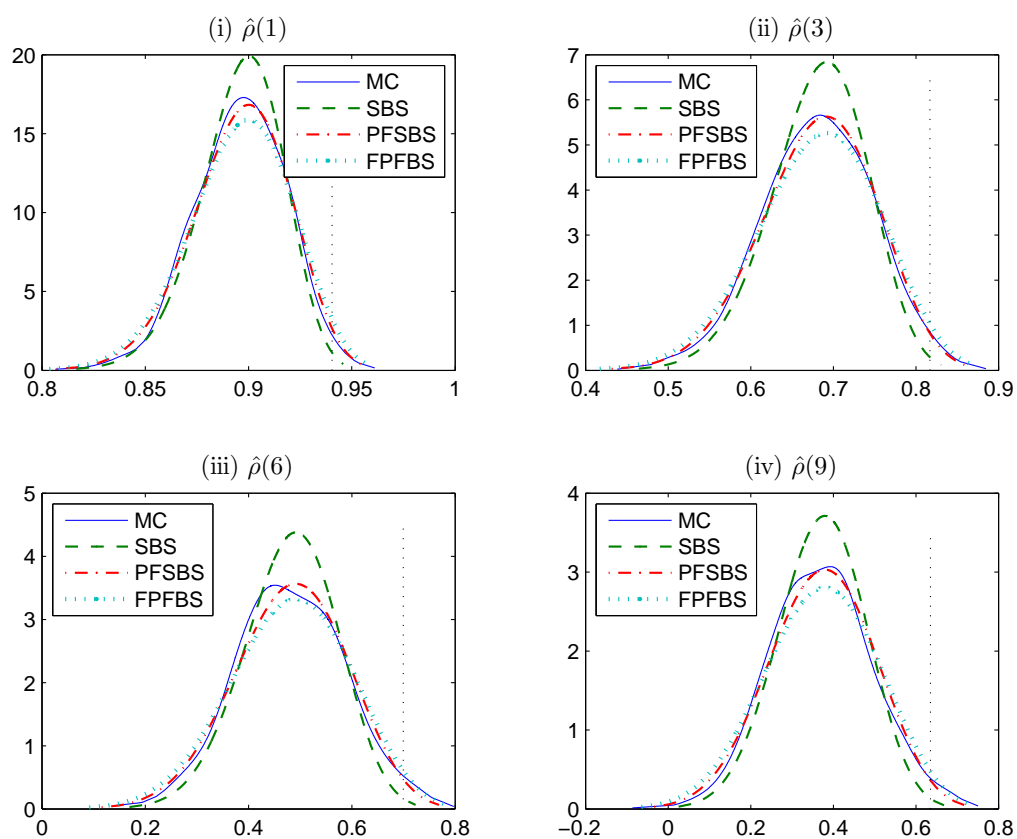


Figure 10: Densities of the sample autocorrelation coefficient $\hat{\rho}(k)$, $k = 1, 3, 6, 9$: under ARFIMA(1, d , 0) with $T = 500$, $d = 0.4$, $\phi = 0.6$: Monte Carlo (MC); averaged (raw) sieve bootstrap (SBS); averaged pre-filtered sieve bootstrap (PFSBS); averaged fixed pre-filtered sieve bootstrap (FPFBS). The vertical dotted line indicates the position of the true value of $\rho(k)$, $k = 1, 3, 6, 9$.

6 Summary and Conclusion

This paper has derived new results regarding the convergence rates of sieve-based bootstrap techniques, in the context of fractionally integrated processes. Both the raw sieve technique, based on an autoregressive approximation of the long memory process, and a pre-filtered version of the sieve method, are investigated, for a broad class of statistics that includes the sample mean and sample second-order moments. Pre-filtering via an appropriate estimator is shown to yield a convergence rate that is equivalent to that associated with intermediate and short memory processes, which is, in turn, arbitrarily close to that associated with independent data.

Using numerical simulation, the distinct (and only rarely noted) problem of underestimating the sampling variance of the sample mean in the long memory case is shown to be avoided, in large measure, by use of a pre-filtering method based, in turn, on a (bias-adjusted) semi-parametric estimator of the long memory parameter. In particular, for moderate values of d the pre-filtered sieve produces very accurate estimates of the (known) exact distribution of the sample mean, and achieves reasonable accuracy elsewhere in the parameter space. Replacing the data-based pre-filter with a fixed value may produce a slight improvement, but only when the latter is close to the true parameter; otherwise the fixed pre-filtering performs very badly in terms of reproducing the exact distribution.

The (data-based) pre-filtering technique is also shown to produce very accurate estimates of the true sampling distribution of selected autocorrelation coefficients (as measured by Monte Carlo simulation). Whilst there is no clear cut superiority of the pre-filtered over the raw sieve method when the fractional integration parameter is small, as the fractional integration parameter increases the performance of the two methods becomes more distinct, and the pre-filtering method performs notably better, reflecting the properties established in the theoretical development. As is the case with the sample mean, while fixed pre-filtering can outperform the raw method when the assigned pre-filtering value is close to the true parameter, it does very poorly otherwise, and in any case never outperforms data-based pre-filtering in terms of reproducing the (Monte Carlo) sampling distributions of the sample autocorrelations. Finally, for the narrow region of the parameter space in which an Edgeworth approximation of the distribution of the sample autocorrelations is valid, the sieve bootstrap reproduces this analytical result with great accuracy.

With due acknowledgement made of the limited nature of the current experimental exercise, we conclude that the overall increase in accuracy obtained when using the data-based pre-filtered sieve bootstrap in parts of the parameter space associated with moderate to strong long-range dependence is worth the slight reduction that might occur (relative to the raw sieve) otherwise; and that a reasonable approach to estimating unknown sampling distributions in empirical settings is to employ the data-based

pre-filtered sieve as the default method.

Appendix: Edgeworth expansion for the sample autocorrelation function

To support the reproducibility of the results reported in this paper, we provide a brief outline of the details of the Edgeworth expansion used as a comparator of our bootstrap-based methodology. All further details of this expansion can be found in [Lieberman et al. \(2001\)](#) (LRZ hereafter).

Suppose we possess a statistic s_T such that $\zeta_T = \sqrt{T}(s_T - E[s_T]) \xrightarrow{D} \mathbb{N}(0, \nu^2)$. The conventional (second-order) Edgeworth expansion for the CDF of ζ_T is of the form

$$\begin{aligned} F_{\zeta_T}(c) &= \Pr\{\zeta_T < c = \nu u\} \\ &= G(u) - \left\{ \frac{\eta_3}{6\sqrt{T}} H_2(u) + \frac{1}{T} \left(\frac{\eta_4}{24} H_3(u) + \frac{\eta_3^2}{72} H_5(u) \right) \right\} g(u) + O(T^{-3/2}) \end{aligned} \quad (\text{A.1})$$

where $\eta_r = \kappa_r / \nu^r$ denotes the r^{th} standardised cumulant of ζ_T , and

$$\begin{aligned} H_2(u) &= u^2 - 1 \\ H_3(u) &= u^3 - 3u \\ H_5(u) &= u^5 - 10u^3 + 15u \end{aligned}$$

are the required Hermite polynomials ([Hall, 1992](#)). Accordingly, direct application of [\(A.1\)](#) to the statistic

$$W_T(k) = \sqrt{T}(\hat{\rho}(k) - \rho(k)) \quad (\text{A.2})$$

requires a means of computing the required cumulants of W . In theory these might be computed via [Magnus \(1986, Theorem 6\)](#), or possibly [Smith \(1989\)](#); in practice these expressions quickly become unmanageable as the order of the required moments increases.

LRZ instead begin with

$$Z_T(k) = \sqrt{n}(\hat{\gamma}(k) - \gamma(k)), \quad (\text{A.3})$$

where

$$\begin{aligned} \hat{\gamma}(k) &= \frac{1}{T} \sum_{t=1}^{T-k} x(t)x(T+k) = T^{-1}x' A_{T,k}x, \\ [A_{T,k}]_{i,j} &= \xi_k(i-j) = \begin{cases} \frac{1}{2} & \text{for } |i-j| = k \\ 0 & \text{otherwise} \end{cases}, \end{aligned}$$

and $A_{T,0} = I_T$. Assuming that $x = (x(1), \dots, x(T))'$ is distributed $\mathbb{N}(0, \Sigma)$ ¹⁰, LRZ then proceed to produce an expansion for $W_T(k)$ indirectly, via $Z_T(k)$, as follows.

For brevity write $Z_T(k)$ and $W_T(k)$ as Z_k and W_k respectively, and use $\rho(k) = \gamma(k)/\gamma(0)$ and $\hat{\gamma}(k) = \gamma(k) + T^{-1/2}Z_k$ to now rewrite W_k as

$$\begin{aligned} W_k &= \sqrt{T} \left(\frac{\hat{\gamma}(k)}{\hat{\gamma}(0)} - \frac{\gamma(k)}{\gamma(0)} \right) \\ &= \sqrt{T} \left(\frac{\hat{\gamma}(k)\gamma(0) - \gamma(k)\hat{\gamma}(0)}{\hat{\gamma}(0)\gamma(0)} \right) \\ &= \frac{\gamma(0)Z_k - \gamma(k)Z_0}{\gamma(0)\hat{\gamma}(0)} \\ &= \frac{Z_k - \rho(k)Z_0}{\hat{\gamma}(0)}. \end{aligned}$$

Then for a single W_k :

$$\begin{aligned} F_{W_k}(c) &= \Pr(W_k < c) \\ &= \Pr(Z_k - \rho(k)Z_0 < c\hat{\gamma}(0)) \quad \text{since } \hat{\gamma}(0) > 0 \\ &= \Pr(Z_k - \rho(k)Z_0 < c\gamma(0) + cT^{-1/2}Z_0) \\ &= \Pr(Z_k - (\rho(k) + cT^{-1/2})Z_0 < c\gamma(0)), \end{aligned}$$

where, from (A.3),

$$\begin{aligned} Z_k - (\rho(k) + cT^{-1/2})Z_0 &= \sqrt{T}(\hat{\gamma}(k) - (\rho(k) + cT^{-1/2})\hat{\gamma}(0)) \\ &\quad - \left[\sqrt{T}(\gamma(k) - (\rho(k) + cT^{-1/2})\gamma(0)) \right] \\ &= \sqrt{T}(\hat{\gamma}(k) - (\rho(k) + cT^{-1/2})\hat{\gamma}(0)) + c\gamma(0) \\ &= \sqrt{T}(n^{-1}x'A_{T,k}x - (\rho(k) + cT^{-1/2})T^{-1}x'x) + c\gamma(0) \\ &= T^{-1/2}x'B_{T,k}x + c\gamma(0) \end{aligned}$$

and $B_{T,k} = A_{T,k} - (\rho(k) + cT^{-1/2})I_T$. So, defining $x'B_{T,k}x = Q_{T,k}^\dagger$, we have

$$F_{W_k}(c) = \Pr(W_k < c) \equiv \Pr\left\{Q_{T,k}^\dagger < 0\right\}.$$

Standard results on quadratic forms in normal variates when $x \sim \mathbb{N}(0, \Sigma)$ gives $\psi_T(\tau) = E[i\tau Q_{T,k}^\dagger] = \prod_{t=1}^T (1 - 2i\tau\lambda_t)^{-\frac{1}{2}}$ on application of Aitken's integral, where λ_t , $t = 1, \dots, T$ are the eigenvalues of $B_{T,k}\Sigma$. The characteristic function is integrable for all $T > 2$ and the cumulant generating function $-\frac{1}{2} \sum_{t=1}^T \log(1 - 2i\tau\lambda_t)$ yields the r^{th} cumulant of $Q_{T,k}^\dagger$ as

$$\kappa_r^* = 2^{r-1} (r-1)! \text{tr} [(B_{T,k}\Sigma)^r].$$

¹⁰ LRZ explicitly impose $E(X) = 0$; or, equivalently, assume that $X = Y - \mu$ where $\mu = E(Y)$ is known.

Evaluating the mean and variance now makes $Q_{T,k}^\dagger$ (or more correctly its z -score) a convenient candidate for an Edgeworth expansion.

From the preceding,

$$F_{W_k}(c) \equiv \Pr \left\{ Q_{T,k}^\dagger < 0 \right\} = \Pr \left\{ \frac{Q_{T,k}^\dagger - \mu^\dagger}{\sigma^\dagger} < u = -\frac{\mu^\dagger}{\sigma^\dagger} \right\},$$

where $\mu^\dagger = \kappa_1^\dagger$ and $\sigma^\dagger = \sqrt{\kappa_2^\dagger}$. Hence the second-order Edgeworth expansion (if it exists) for the CDF of $Q_{T,k}^\dagger$ (and hence W_k) will be of the form¹¹

$$\tilde{F}_{W_k}(c) = G(u) - \left\{ \frac{\eta_3^\dagger}{6} H_2(u) + \frac{\eta_4^\dagger}{24} H_3(u) + \frac{(\eta_3^\dagger)^2}{72} H_5(u) \right\} g(u),$$

with error $O(T^{-3/2})$, where $\eta_r^\dagger = \kappa_r^\dagger / (\sqrt{\kappa_2^\dagger})^r$, $r = 1, 2, 3, 4$, and $u = -\eta_1^\dagger$. Note that the descending powers of \sqrt{T} that would ordinarily appear in the expansion (*cf.* (A.1)) are here subsumed into the standardised cumulants; that is, we are implicitly assuming that $\eta_r^\dagger = O(T^{1-r/2})$ or, equivalently, that κ_r^\dagger is $O(T)$, at least up to $r = 4$.

That the cumulants of $Q_{T,k}^\dagger$ are of the appropriate order, at least for restricted values of the fractional parameter d , follows from LRZ Theorem 1. In particular, the cumulants of $Q_{T,k}^\dagger$ of order no greater than r will be $O(T)$ only if $r(2d) < 1$, implying that κ_r^\dagger , $r = 1, 2, 3, 4$, are $O(T)$ if $d < 0.125$ but not otherwise. However, if r now denotes the order of the highest cumulant in the expansion, then LRZ also show that we require $(r+1)(2d) < 1$ and r even to attain an expansion error of $o(T^{1-r/2})$; while if r is odd the error is of the same order as the last term, namely $O(T^{1-r/2})$. Hence the second-order ($r = 4$) expansion is valid only for $d < 0.1$, and there is no valid expansion (in the sense that the error is of smaller order than the last included term) for $d \geq 0.1$.

References

- ADENSTEDT, R. K. (1974). On large-sample estimation for the mean of a stationary random sequence. *The Annals of Statistics*, **2** pp. 1095–1107. URL <http://www.jstor.org/stable/2958332>.
- ANDREWS, D. W., LIEBERMAN, O. and MARMER, V. (2006). Higher-order improvements of the parametric bootstrap for long-memory Gaussian processes. *Journal of Econometrics*, **133** 673–702.
- ANDREWS, D. W. K. and LIEBERMAN, O. (2005). Valid Edgeworth expansions

¹¹ It should be noted that LRZ give the expansion for $1 - F_{W_k}(c)$, rather than $F_{W_k}(c)$ itself.

- for the Whittle maximum likelihood estimator for stationary long-memory gaussian time series. *Econometric Theory*, **21** 710–734.
- ANDREWS, D. W. K. and SUN, Y. (2004). Adaptive local polynomial Whittle estimation of long-range dependence. *Econometrica*, **72** 569–614.
- APOSTOL, T. M. (1960). *Mathematical Analysis*. Addison-Wesley, Reading.
- BERAN, J. (1994). *Statistics for long-memory processes*, vol. 61 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, New York.
- BERAN, J. (1995). Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models. *Journal of the Royal Statistical Society*, **B 57** 654–672.
- BICKEL, P. J. and FREEDMAN, D. A. (1981). Some asymptotic theory for the bootstrap. *Annals of Statistics*, **9** 1196–1217.
- BROCKWELL, P. J. and DAVIS, R. A. (1991). *Time Series: Theory and Methods*. 2nd ed. Springer Series in Statistics, Springer-Verlag, New York.
- BÜHLMANN, P. (1997). Sieve bootstrap for time series. *Bernoulli*, **3** 123–148.
- CHOI, E. and HALL, P. G. (2000). Bootstrap confidence regions from autoregressions of arbitrary order. *Journal of the Royal Statistical Society*, **B 62** 461–477.
- DAHLHAUS, R. (1989). Efficient parameter estimation for self-similar processes. *Annals of Statistics*, **17** 1749–1766.
- DOORNIK, J. A. and OOMS, M. (2001). Computational aspects of maximum likelihood estimation of autoregressive fractionally integrated moving average models. *Computational Statistics & Data Analysis*, **42** 333–348. Also a 2001 Nuffield discussion paper.
- DURBIN, J. (1980). Approximations for densities of sufficient estimators. *Biometrika*, **67** 311–333.
- FOX, R. and TAQQU, M. S. (1986). Large sample properties of parameter estimates for strongly dependent stationary gaussian time series. *Annals of Statistics*, **14** 517–532.
- GEWEKE, J. and PORTER-HUDAK, S. (1983). The estimation and application of long-memory time series models. *Journal of Time Series Analysis*, **4** 221–238.
- GIRAITIS, L. and ROBINSON, P. M. (2003). Edgeworth expansions for semiparametric Whittle estimation of long memory. *Annals of Statistics*, **31** 1325–1375.

- GRANGER, C. W. J. and JOYEUX, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, **1** 15–29.
- HALL, P. (1992). *The bootstrap and Edgeworth expansion*. Springer series in Statistics, Springer-Verlag, New York.
- HESTERBERG, T. (1997). Matched-block bootstrap for long memory processes. Research Report 66, MathSoft, Inc, Seattle, WA.
- HOSKING, J. R. M. (1980). Fractional differencing. *Biometrika*, **68** 165–176.
- HOSKING, J. R. M. (1996). Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long memory time series. *Journal of Econometrics*, **73** 261–284.
- INOUE, A. and KASAHARA, Y. (2006). Explicit representation of finite predictor coefficients and its applications. *Annals of Statistics*, **34** 973–993.
- KREISS, J. P., PAPANODITIS, E. and POLITIS, D. N. (2011). On the range of validity of the autoregressive sieve bootstrap. *Annals of Statistics*, **39** 2103–2130.
- KÜNSCH, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics*, **17** 1217–1241.
- LIEBERMAN, O. and PHILLIPS, P. C. B. (2004). Expansions for the distribution of the maximum likelihood estimator of the fractional difference parameter. *Econometric Theory*, **20** 464–484.
- LIEBERMAN, O., ROUSSEAU, J. and ZUCKER, D. M. (2001). Valid Edgeworth expansion for the sample autocorrelation function under long range dependence. *Econometric Theory*, **17** 257–275.
- LIEBERMAN, O., ROUSSEAU, J. and ZUCKER, D. M. (2003). Valid asymptotic expansions for the maximum likelihood estimator of the parameter of a stationary, gaussian, strongly dependent process. *The Annals of Statistics*, **31** 586–612.
- MAGNUS, J. R. (1986). The exact moments of a ratio of quadratic forms in normal variables. *Annals of Economics and Statistics / Annales d'Économie et de Statistique* pp. 95–109. URL <http://www.jstor.org/stable/20075629>.
- NIELSEN, M. Ø. and FREDERIKSEN, P. H. (2005). Finite sample comparison of parametric, semiparametric, and wavelet estimators of fractional integration. *Econometric Reviews*, **24** 405–443.
- POLITIS, D. N. (2003). The impact of bootstrap methods on time series analysis. *Statistical Science*, **18** 219–230.

- POSKITT, D. S. (1994). A note on autoregressive modelling. *Econometric Theory*, **10** 884–899.
- POSKITT, D. S. (2007). Autoregressive approximation in nonstandard situations: The fractionally integrated and non-invertible cases. *Annals of Institute of Statistical Mathematics*, **59** 697–725.
- POSKITT, D. S. (2008). Properties of the sieve bootstrap for fractionally integrated and non-invertible processes. *Journal of Time Series Analysis*, **29** 224–250.
- POSKITT, D. S., MARTIN, G. M. and GROSE, S. G. (2012). Bias reduction of long memory parameter estimators via the pre-filtered sieve bootstrap. Econometrics & Business Statistics Working Paper WP 08/12, Monash University.
- ROBINSON, P. M. (1995a). Gaussian semiparametric estimation of long range dependence. *Annals of Statistics*, **23** 1630–1661.
- ROBINSON, P. M. (1995b). Log periodogram regression of time series with long memory. *Annals of Statistics*, **23** 1048–1072.
- 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.
- SMITH, M. D. (1989). On the expectation of a ratio of quadratic forms in normal variables. *Journal of Multivariate Analysis*, **31** 244–257. URL <http://EconPapers.repec.org/RePEc:eee:jmvana:v:31:y:1989:i:2:p:244-257>.
- SOWELL, F. (1992). Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics*, **53** 165–188.
- TANIGUCHI, M. (1984). Validity of Edgeworth expansions for statistics of time series. *Journal of Time Series Analysis*, **5** 37–51.