ISSN 1440-771X

MONASH University

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

April 2012

Working Paper 09/12

Higher Order Improvements of the Sieve Bootstrap for Fractionally Integrated Processes

D. S. Poskitt, Simone D. Grose & Gael M. Martin *

Department of Econometrics & Business Statistics, Monash University

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 linear statistics. The methods are then applied to the problem of estimating the sampling distribution of the sample mean under long memory, in an experimental setting. 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.

Keywords: Bias, bootstrap-based inference, Edgeworth expansion, pre-filtered sieve bootstrap, 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

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

*Corresponding author: Gael Martin, Department of Econometrics and Business Statistics, Monash University, Victoria 3800, Australia Tel.:+61-3-9905-1189; fax:+61-3-9905-5474. *E-mail address:* gael.martin@monash.edu. This research has been supported by Australian Research Council (ARC) Discovery Grant DP0985234 and ARC Future Fellowship FT0991045.

where $\varepsilon(t)$, $t \in \mathcal{Z}$, 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 Wold representation (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 Taqqu (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 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. As such, attention has also been given to the application of bootstrap-based inference. 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 ARFIMA(p, d, q) models. In contrast, Poskitt (2008) proposes a non-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 "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 – what we will refer to as – the raw sieve bootstrap can achieve an error rate of $O(T^{-(1-d')+\beta})$ for all $\beta > 0$ where $d' = \max\{0, d\}$. 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 will 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 error rate of the pre–filtered sieve bootstrap is $O(T^{-1+\beta})$ for all $\beta > 0$. These results generalize those of Choi and Hall (2000) who show in the short–memory case that for double bootstrap calibrated percentile methods, and percentile t confidence intervals, the convergence rate of the sieve bootstrap is arbitrarily close to that obtained with simple random samples, namely $O(T^{-1+\beta})$ 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), and our results suggest, a–fortiori, that the same is true with fractionally integrated processes.

We illustrate our results by means of a small simulation study, in which we examine the sieve bootstrap approximation to the sampling distribution of the sample mean. We compare and contrast the performance of the raw and the pre–filtered sieve bootstrap in correctly characterizing the known properties of the sample mean under long memory, most notably its finite sample variance.

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 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 $t \in \mathcal{Z}$ denote a linearly regular, covariance-stationary process with Wold representation as in (1) where the innovations and the impulse response coefficients satisfy the following conditions:

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

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

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

Assumption 2. *The transfer function in the Wold representation of the process $y(t)$, namely $k(z) = \sum_{j\geq 0} k(j)z^j$, 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 implies the innovations are fundamental and incorporates 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

$$\bar{y}_h(t) = \sum_{j=1}^{h} \pi_h(j) y(t-j) \equiv -\sum_{j=1}^{h} \phi_h(j) y(t-j); \tag{3}$$

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

$$\varepsilon_h(t) = \sum_{j=0}^{h} \phi_h(j) y(t-j). \tag{4}$$

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

$$\sum_{j=0}^{h} \phi_h(j) \gamma(j-k) = \delta_0(k) \sigma_h^2\,, \quad k = 0, 1, \ldots, h, \tag{5}$$

in which $\gamma(\tau) = \gamma(-\tau) = E[y(t)y(t-\tau)]$, $\tau = 0, 1, \ldots$ 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

$$\sigma_h^2 = E\big[\varepsilon_h(t)^2\big] \tag{6}$$

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

The use of finite-order AR 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 be a good approximation to the "infinite-order" predictor $\bar{y}(t)$ for sufficiently large h.

The asymptotic validity, and properties, of finite order autoregressive models when $h \to \infty$ with the sample size T under regularity conditions which admit non-summable processes was proved 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 approximating autoregression 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 Grose and Poskitt (2006). The sieve bootstrap of Bühlmann (1997), which works by "whitening" the data using an autoregressive approximation, with the dynamics of the process captured in the fitted autoregression, is accordingly a plausible non-parametric bootstrap technique for long-memory processes.

Details of the sieve bootstrap for fractional processes are given in Poskitt (2008). Here we present for convenience the basic steps needed to generate a sieve bootstrap realization of a process $y(t)$ (referred to as the sieve bootstrap (SB) algorithm hereafter):

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

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

using $y(1-j) = y(T-j+1)$, $j = 1, \ldots, h$, as initial values. From $\bar{\varepsilon}_h(t)$, $t = 1, \ldots, T$, 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 $\varepsilon_h^+(t)$, $t = 1, \ldots, T$, 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 $\tilde{\varepsilon}_h(t)$, $t = 1, \ldots, T$. Set $\varepsilon_h^*(t) = \bar{\sigma}_h \varepsilon_h^+(t)$, $t = 1, \ldots, T$.

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

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

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

Crucially, in the fractional case the rate of convergence of the coefficient estimates $\bar{\phi}_h(1), \ldots, \bar{\phi}_h(h)$ 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* $h \leq H_T = a(\log T)^c$, $a > 0$, $c < \infty$,

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

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). □

Consider a statistic $\mathbf{s}_T = (s_{1T}, \ldots, s_{mT})'$ where $s_{iT} = s_i(y(1), \ldots, y(T))$ and each $s_i(\cdot)$ for $i = 1, \ldots, m$ is a suitably smooth function of the time series values $y(1), \ldots, y(T)$. 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 $y^*(1), \ldots, y^*(T)$, a realization obtained from the SB algorithm, so that $\mathbf{s}_T^* = (s_{1T}^*, \ldots, s_{mT}^*)'$ where $s_{iT}^* = s_i(y^*(1), \ldots, y^*(T))$. 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 $b = 1, \ldots, B$, we can approximate $F_{\mathbf{s}_T^*}(\mathbf{s})$ by the bootstrap empirical distribution function

$$\bar{F}_{\mathbf{s}_T^*,B}(\mathbf{s}) = B^{-1}\sum_{b=1}^{B}\mathbf{1}\{\mathbf{s}_{T,b}^* \leq \mathbf{s}\}\,. \tag{7}$$

By the (strong) Glivenko-Cantelli Theorem

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

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 $F_{\mathbf{s}_T}(\mathbf{s})$.

For the sieve bootstrap Poskitt (2008) shows that under Assumptions 1 and 2, so called, linear statistics satisfy $\eta(F_{\mathbf{s}_T^*}, F_{\mathbf{s}_T}) = o(T^{-1/2(1-2d')+\beta})$ for all $\beta > 0$ wherein $d' = \max\{0, d\}$. Here $\eta(F_X, F_Y)$ denotes Mallow's measure of the distance between two probability distributions

F_X and F_Y, see Bickel and Freedman (1981, Section 8) for a discussion of Mallow's metric and its properties. The class of linear statistics considered in Poskitt (2008) is the same as that considered in Künsch (1989, Section 2.1) and Bühlmann (1997, Section 3.3). But Mallows metric is equivalent to weak convergence (Bickel and Freedman, 1981, Lemma 8.3), and in conjunction with a $o(T^{-1/2(1-2d')+\beta})$ convergence rate this intimates that use of the sieve bootstrap may be little better than using central limit properties. However, in what follows we show that for a more restricted range of statistics this 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

In order to show that the convergence rate given in Poskitt (2008) can be improved upon 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 constant $M < \infty$ such that*

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

Various statistics, such as the sample mean, for example, the sample autocovariances, autocorrelations and partial-autocorrelations, satisfy Assumption 4.

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

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

hold where $s = [m/2] + 1$ and $r \geq 3$, and secondly, $\partial^q \log\psi_T(\boldsymbol{\tau})/\partial\boldsymbol{\tau}^q$ exists for all $\boldsymbol{\tau}$ in a neighbourhood of the origin and $\lim_{\|\boldsymbol{\tau}\|\to 0} T^{-1}\partial^q \log\psi_T(\boldsymbol{\tau})/\partial\boldsymbol{\tau}^q$ exists as $T \to \infty$ for all $q = 1, \ldots, q' = \max\{s, r+1\}$.

Here E denotes the 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).

Let $\mathbf{V}_T = T^{-1}E\left[(\mathbf{s}_T - E[\mathbf{s}_T])(\mathbf{s}_T - E[\mathbf{s}_T])'\right]$ and set $\boldsymbol{\zeta}_T = \mathbf{V}_T^{-1/2}(\mathbf{s}_T - E[\mathbf{s}_T])$. 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 formal Edgeworth expansion

$$P(\boldsymbol{\zeta}_T \leq \mathbf{z}) = G(\mathbf{z}) + \sum_{j=3}^{r} T^{1-j/2}\pi_j(\mathbf{z}, \mathbf{K}_r)g(\mathbf{z}) + o(T^{1-r/2}) \tag{8}$$

uniformly in $\mathbf{z}$, where $G(\mathbf{z})$ denotes the distribution function of a Gaussian $N(\mathbf{0}, \mathbf{I}_m)$ random vector, $g(\mathbf{z})$ the corresponding density, and $\pi_j(\mathbf{z}, \mathbf{K}_r)$ is a polynomial function of degree j in $\mathbf{z}$ whose coefficients are polynomials in the elements of the cumulants $\mathbf{K}_r = (\mathbf{k}_1', \ldots, \mathbf{k}_r')'$, $\mathbf{k}_r = \imath^{-r}\partial^r \log \psi_T(\mathbf{0})/\partial \boldsymbol{\tau}^r$. See Theorem 1 of Taniguchi (1984) and Durbin (1980).

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

$$P^*(\boldsymbol{\zeta}_T^* \leq \mathbf{z}) = G(\mathbf{z}) + \sum_{j=3}^{r} T^{1-j/2}\pi_j(\mathbf{z}, \mathbf{K}_r^*)g(\mathbf{z}) + o(T^{1-r/2}) \tag{9}$$

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

A comparison of (8) and (9) for $r \geq 4$ now indicates that

$$\sup_{\mathbf{z}} |P^*(\boldsymbol{\zeta}_T^* \leq \mathbf{z}) - P(\boldsymbol{\zeta}_T \leq \mathbf{z})| = T^{-1/2}O(\|\mathbf{K}_r^* - \mathbf{K}_r\|) + o(T^{-1})\,. \tag{10}$$

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 $\|\mathbf{K}_4^* - \mathbf{K}_4\| = o(T^{-1/2}\varrho_T)$ where $\varrho_T = o(T^{1/2})$ then (10) implies that the bootstrap probability will have an error rate of $O(T^{-1}\varrho_T)$. See Choi and Hall (2000, Appendix A.2). This type of argument was used by Andrews et al. (2006) to show that the parametric bootstrap based on the (approximate) Gaussian MLE achieves an error rate of order $T^{-1}\log T$ for a one sided confidence interval. Using this approach we can establish an analogous result for the sieve bootstrap.

Theorem 6. *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*

$$\sup_{\mathbf{z}} |P^*(\boldsymbol{\zeta}_T^* \leq \mathbf{z}) - P(\boldsymbol{\zeta}_T \leq \mathbf{z})| = O(T^{-(1-d')+\beta})$$

for all $\beta > 0$.

The proof of Theorem 6 relies on the following lemma. The heuristics behind the lemma 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 $\beta > 0$, and for all $\boldsymbol{\tau}$ such that $\|\boldsymbol{\tau}\| \leq T^{\beta/2}$ the difference $|\psi_T^*(\boldsymbol{\tau}) - \psi_T(\boldsymbol{\tau})| = o\{T^{-1/2(1-2d')+\beta}\}$ uniformly in $\boldsymbol{\tau}$.*

Proof: Given that $\mathbf{s}_T$ satisfies Assumption 4, $\|\mathbf{s}_T^* - \mathbf{s}_T\|^2 \leq mMT^{-1}\sum_{t=1}^T (y^*(t) - y(t))^2$where $M < \infty$. Arguing as in the proof of (Theorem 4.1 Poskitt, 2008, p.16-18) 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 $\exp(\imath x)$ 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 $|\exp(\imath\boldsymbol{\tau}'\mathbf{x}) - \exp(\imath\boldsymbol{\tau}'\mathbf{y})| \leq KT^{\beta/2}\|\mathbf{x} - \mathbf{y}\|$. Then, as in Bickel and Freedman (1981, p. 1212),

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

But $E[E^*[\|\mathbf{s}_T^* - \mathbf{s}_T\|]] \leq E[E^*[\|\mathbf{s}_T^* - \mathbf{s}_T\|^2]]^{1/2} = o(T^{-1/2(1-2d'-\beta)})$, giving the required result. □

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 $\|\mathbf{K}_q^* - \mathbf{K}_q\| = o(T^{-1/2+d'+\beta})$ for $q = 1, \ldots, q'$ and all $\beta > 0$.*

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 $\log\psi_T^*(\boldsymbol{\tau}) - \log\psi_T(\boldsymbol{\tau}) = (\psi_T^*(\boldsymbol{\tau}) - \psi_T(\boldsymbol{\tau}))/\psi_T(\boldsymbol{\tau}) + O\left(|(\psi_T^*(\boldsymbol{\tau}) - \psi_T(\boldsymbol{\tau}))/\psi_T(\boldsymbol{\tau})|^2\right) = o(T^{-1/2(1-2d')+\beta})$ uniformly in $\boldsymbol{\tau}$ by Lemma 7.

Now set

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

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

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

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

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

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

$$\lim_{T\to\infty}\lim_{\|\mathbf{t}-\boldsymbol{\tau}\|\to 0} |\Delta_T^*(\mathbf{t};\boldsymbol{\tau})| \leq \lim_{\|\mathbf{t}-\boldsymbol{\tau}\|\to 0}\lim_{T\to\infty} \frac{|\log\psi_T^*(\mathbf{t}) - \log\psi_T(\mathbf{t})| + |\log\psi_T^*(\boldsymbol{\tau}) - \log\psi_T(\boldsymbol{\tau})|}{\|\mathbf{t}-\boldsymbol{\tau}\|}$$
$$= o(T^{-1/2(1-2d')+\beta})\,.$$

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

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

where $\mathbf{u}_j = (0,\ldots,0,1,0\ldots,0)'$, the jth 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 $\|\mathbf{K}_1^* - \mathbf{K}_1\| = o(T^{-1/2+d'+\beta})$.

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 $\|\mathbf{K}_{j+1}^* - \mathbf{K}_{j+1}\| = o(T^{-1/2+d'+\beta})$. Induction on $j = 1,\ldots,q'$ completes the proof. □

Proof of Theorem 6: By construction the bootstrap innovations $\varepsilon_h^*(t)$ in Step S2 of the sieve bootstrap satisfy Assumption 1, and the sieve bootstrap process $y^*(t)$ produced in Step S3 satisfies Assumption 2. The statistic $\mathbf{s}_T^*$ therefore satisfies Assumption 4 and Assumption 5 with $r \geq 4$, and Assumption 5 validates the formal Edgeworth expansion in (9) in the same manner that it justifies (8). Corollary 8 implies that $\|\mathbf{K}_4^* - \mathbf{K}_4\| = o(T^{-1/2+d'+\beta})$ and Theorem 6 then follows from equation (10) as previously outlined. □

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

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

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

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

for all $\beta > 0$. This leads to a coverage probability for $S(q_\alpha^*)$ of $(1-\alpha) + O(T^{-(1-d)+\beta})$ for all $\beta > 0$ when $d \in (0, 0.5)$, the long memory case, compared to $(1-\alpha) + O(T^{-1+\beta})$ 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. Obviously, in the long memory case the closer is d to zero the closer the convergence rate will be to the rate achieved with short memory and anti-persistent processes. 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 value $\widehat{d}$ is available such that $\widehat{d} - d \in N_\delta = \{x : |x| < \delta\}$ where $2\delta < \min\{|d|, 0.5 - |d|\}$. For any $d > -1$ let $\alpha_j^{(d)}$, $j = 0, 1, 2, \ldots$, denote the coefficients of the fractional difference operator when expressed in terms of its binomial expansion,

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

and set

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

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

PFSB1. Calculate the coefficients of the filter $(1-z)^{\widehat{d}}$ and from the data generate the filtered values

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

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

PFSB2. Fit an AR approximation to $\widehat{w}(t)$ and generate a sieve bootstrap sample $\widehat{w}^*(t)$, $t = 1, \ldots, T$, of the filtered data as in Steps SB1–SB3 of the SB algorithm.

PFSB3. Using the coefficients of the (inverse) filter $(1-z)^{-\widehat{d}}$ construct a corresponding pre–filtered sieve bootstrap draw

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

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

We will refer to this as the PFSB algorithm.

Note that the process

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

has fractional index $d-\widehat{d}$. By assumption $|d-\widehat{d}|<\delta$ and the error in the AR approximation to $\widehat{w}(t)$ 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 $\widehat{d}-d\in N_\delta$ where $2\delta<\min\{|d|,0.5-|d|\}$. 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

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

from which it follows that

$$\widehat{y}^*(t)-y(t)=\sum_{j=0}^{t-1}\alpha_j^{(-\widehat{d})}\{\widehat{w}^*(t-j)-\widehat{w}(t-j)\} \tag{11}$$

for all $t=1,\ldots,T$.

Now let $Z_{\{\widehat{w}^*-\widehat{w}\}}(\lambda)$, $\lambda\in[-\pi,\pi]$, denote the process of orthogonal increments associated with $\widehat{w}^*(t)-\widehat{w}(t)$; which we take as the mean square limit as $T\to\infty$ of

$$\frac{1}{2\pi}\sum_{t=1}^{T}\frac{e^{-\imath\lambda t}-1}{-\imath t}\{\widehat{w}^*(t)-\widehat{w}(t)\}\,,$$

in the product space $(\Omega\otimes\Omega^*,\mathfrak{F}\otimes\mathfrak{F}^*,P(P^*))$ with joint distribution corresponding to the marginal and conditional probability measures P and P^*. Then, in differential notation,

$$E[E^*[dZ_{\{\widehat{w}^*-\widehat{w}\}}(\lambda)\overline{dZ_{\{\widehat{w}^*-\widehat{w}\}}(\nu)}]]=\delta(\lambda-\nu)dF_{\{\widehat{w}^*-\widehat{w}\}}(\lambda)d\nu$$

where $F_{\{\widehat{w}^*-\widehat{w}\}}(\lambda)$ is the spectral distribution of $\widehat{w}^*(t)-\widehat{w}(t)$.

Reexpressing equation (11) using the Cramèr representation and employing the integral version of the Cauchy-Schwartz inequality gives

$$\begin{aligned}|\widehat{y}^*(t) - y(t)| &= \left| \int_{-\pi}^{\pi} e^{\imath\lambda t} \left(\sum_{j=0}^{t-1} \alpha_j^{(-\widehat{d})} e^{-\imath\lambda j} \right) dZ_{\{\widehat{w}^* - \widehat{w}\}}(\lambda) \right| \\ &\leq \left(\int_{-\pi}^{\pi} \left| \sum_{j=0}^{t-1} \alpha_j^{(-\widehat{d})} e^{-\imath\lambda j} \right|^2 d\lambda \right)^{1/2} \cdot \left(\int_{-\pi}^{\pi} |dZ_{\{\widehat{w}^* - \widehat{w}\}}(\lambda)|^2 \right)^{1/2}\end{aligned}$$

with probability one. From Parseval's identity we have

$$\int_{-\pi}^{\pi} \left| \sum_{j=0}^{t-1} \alpha_j^{(-\widehat{d})} e^{-\imath\lambda j} \right|^2 d\lambda = 2\pi \sum_{j=0}^{t-1} |\alpha_j^{(-\widehat{d})}|^2 \leq 2\pi \sum_{j=0}^{\infty} |\alpha_j^{(-\widehat{d})}|^2 = \frac{2\pi\Gamma(1-2\widehat{d})}{(\Gamma(1-\widehat{d}))^2} < \frac{2\pi\Gamma(\delta)}{(\Gamma(0.5+\delta))^2} .$$

Using the well known isometric isomorphism between the time and frequency domains we can therefore conclude that

$$\begin{aligned}E[E^*[(\widehat{y}^*(t) - y(t))^2]] &\leq \frac{2\pi\Gamma(\delta)}{(\Gamma(0.5+\delta))^2} \int_{-\pi}^{\pi} dF_{\{\widehat{w}^* - \widehat{w}\}}(\lambda) \\ &= GE[E^*[(\widehat{w}^*(t) - \widehat{w}(t))^2]] . \quad \square\end{aligned}$$

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

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 $\delta_T < \delta$ and $2\delta < \min\{|d|, 0.5 - |d|\}$, $E[E^*[\|\widehat{\mathbf{s}}_T^* - \mathbf{s}_T\|^2]] = \exp(2\delta_T \log T) o(T^{-1+\beta})$ for all $\beta > 0$. Furthermore, for all $\boldsymbol{\tau}$ such that $\|\boldsymbol{\tau}\| \leq T^{\beta/2}$ the difference $|\widehat{\psi}_T^*(\boldsymbol{\tau}) - \psi_T(\boldsymbol{\tau})| = \exp(\delta_T \log T) o\{T^{-1/2+\beta}\}$ 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 $E[E^*[\|\widehat{\mathbf{s}}_T^* - \mathbf{s}_T\|^2]] \leq mME[E^*[|(\widehat{y}^*(t) - y(t)|^2]]$. But by Proposition 9 $E[E^*[|(\widehat{y}^*(t) - y(t)|^2]] \leq GE[E^*[|(\widehat{w}^*(t) - \widehat{w}(t)|^2]]$ where $G < \infty$ and the argument used in the proof of (Theorem 4.1 Poskitt, 2008, p.16-18) shows that $E[E^*[|(\widehat{w}^*(t) - \widehat{w}(t)|^2]] = o(T^{-(1-2\widehat{\delta}')+\beta})$ where $\widehat{\delta}' = \max\{0, d - \widehat{d}\} < \delta_T$ 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 follows the same line of argument as that used in Lemma 7: For all $\boldsymbol{\tau}$ such that $\|\boldsymbol{\tau}\| \leq T^{\beta/2}$ there exists a $K < \infty$ such that $|\exp(\imath\boldsymbol{\tau}'\mathbf{x}) - \exp(\imath\boldsymbol{\tau}'\mathbf{y})| \leq KT^{\beta/2}\|\mathbf{x}-\mathbf{y}\|$ and hence

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

This gives us the desired conclusion because $E[E^*[\|\widehat{\mathbf{s}}^*_T - \mathbf{s}_T\|]] \leq E[E^*[\|\widehat{\mathbf{s}}^*_T - \mathbf{s}_T\|^2]]^{1/2} = \exp(\delta_T \log T)o\{T^{-1/2(1-\beta)}\}$. □

Now set $\widehat{\boldsymbol{\zeta}}^*_T = \widehat{\mathbf{V}}^{*-1/2}_T(\widehat{\mathbf{s}}^*_T - E^*[\widehat{\mathbf{s}}^*_T])$ where $\widehat{\mathbf{V}}^*_T = T^{-1}E^*\left[(\widehat{\mathbf{s}}^*_T - E^*[\widehat{\mathbf{s}}^*_T])(\widehat{\mathbf{s}}^*_T - E^*[\widehat{\mathbf{s}}^*_T])'\right]$. Lemma 10 implies that $E^*[\widehat{\mathbf{s}}^*_T] = E[\mathbf{s}_T] + o(1)$ and $\widehat{\mathbf{V}}^*_T = \mathbf{V}_T + o(1)$, and leads on to the following result.

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 $\delta_T < \delta$ and $2\delta < \min\{|d|, 0.5 - |d|\}$,*

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

for all $\beta > 0$.

Proof: Apart from minor notational changes and an allowance for the filtering that occurs at Steps FS1 and FS3, the argument leading from Lemma 10 to Theorem 11 is almost identical to that leading from Lemma 7 to Theorem 6. The details are omitted. □

In practice, of course, the preliminary estimate of d will be constructed from the data and from Theorem 11 we can see that if $\delta_T \log T \to 0$ as $T \to \infty$ then the convergence rate of the pre–filtered sieve bootstrap will be $O(T^{-1+\beta})$ for all $\beta > 0$, arbitrarily close to the rate achieved with simple random samples. It follows that we will require the estimator $\widehat{d}$ to be such that $|\widehat{d} - d|\log T = o(1)$.

Suppose that $\widehat{d}$ is a $\sqrt{N}$–CAN estimator of d, that is, $\widehat{d}$ is consistent, $|\widehat{d} - d| = o(1)$, and asymptotically normal, $N^{1/2}(\widehat{d} - d) \xrightarrow{\mathcal{D}} N(0, \upsilon)$, where N is a monotonically increasing function of T such that $N/T \to 0$ as $T \to \infty$. For any ϵ, $0 < \epsilon < 0.5$, it follows from the tail area properties of the normal distribution that $\lim_{T\to\infty} P(|\widehat{d} - d| > N^{-1/2+\epsilon}) \leq \exp(-N^{2\epsilon}/2\upsilon)$. Since $\exp(-N^{\epsilon}/2\upsilon) < |r|^{N^{\epsilon}}$ for all r such that $\exp(-1/2\upsilon) < |r| < 1$ we can conclude from the Borel–Cantelli lemma that $|\widehat{d} - d| = o(N^{-1/2+\epsilon})$. Hence it follows that $\widehat{d}$ will satisfy $|\widehat{d} - d|\log T = o(1)$, as required, if $\log T/N^{1/2-\epsilon} \to 0$.

Clearly, the more accurate the preliminary estimate of d, the more useful the pre–filtering, in terms of yielding a filtered process for which the AR approximation is accurate. Given the non-parametric flavour of our approach, in the simulation exercise that follows we apply a PFBS algorithm based on a pre-filtering value of d that is produced by a combination of analytical and bootstrap-based bias adjustments of a base semi-parametric Gaussian local Whittle (SPWL) estimator. Details of this bias adjustment, including all associated numerical results, are found in Poskitt, Martin and Grose (2012). The simulation design adopted in the current paper is identical to that adopted in Poskitt et al. (2012), so that the bias-adjusted estimator that is found to be 'optimal' for any particular parameter (and sample size) setting therein, can be used here as the pre-filter under the corresponding data generating process.

5 Simulation Exercise

In this section we examine the performance of the sieve bootstrap techniques via a small simulation experiment. Specifically, we investigate the accuracy with which both the raw and pre-filtered sieve algorithms approximate the true sampling distribution of one particular scalar statistic, the sample mean $s_T = \bar{y}_T$. Various properties of $\bar{y}_T$ 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

$$Var[\bar{y}_T] = \frac{1}{T} \sum_{k=1-T}^{T-1} (1 - \frac{|k|}{T}) \gamma(k), \qquad (12)$$

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

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

as $T \to \infty$, and thirdly, that the re–normalized mean $T^{1/2-d}(\bar{y}_T - \mu) \xrightarrow{\mathcal{D}} N(0, \omega^2)$, 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 (8) beyond the first are null. Given knowledge of the true sampling variance of the mean in (12), 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.

5.1 Simulation Design

Data are simulated from a Gaussian $ARFIMA(1, d, 0)$ process,

$$(1 - L)^d \Phi(z) y(t) = \varepsilon(t)\,,\ 0 < d < 0.5\,, \tag{13}$$

with $\Phi(z) = 1 - \phi z$ being the operator for a stationary AR(1) component and $\varepsilon(t)$ is zero-mean Gaussian white noise. The theoretical ACF for this process can be computed using the procedures of Sowell (1992). The process in (13) is simulated $R = 1000$ times for $d = 0.0, 0.2, 0.3, 0.4$; $\phi = 0.3, 0.6, 0.9$, and sample sizes $T = 100, 200, 500$ via Levinson recursion applied to the autocorrelation function (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 sample mean, $\bar{y}_{T,r}$, plus $B = 1000$ bootstrap estimates $\bar{y}^*_{T,r(b)}$, constructed using $b = 1, \ldots, B$ bootstrap re-samples obtained via either the raw SB or the PFSB algorithm. Each realized value $\bar{y}_{T,r}$ thus has associated with it a 'bootstrap distribution' based on the B bootstrap resamples $\bar{y}^*_{T,r(b)}$, $b = 1, \ldots, B$, with each such distribution serving as an estimate of $f(\bar{y}_T)$. In order to compare the R bootstrap distributions with the true Gaussian distribution of $\bar{y}_T$, we first compute an 'average' bootstrap distribution by sorting the B bootstrap draws for each MC replication into ascending order, then average 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-based estimate, when plotted against the true distribution of $\bar{y}_T$, allows a direct visual comparison of the overall location and variation of the bootstrap distribution with the actual distribution of $\bar{y}_T$. We also tabulate the ratio of the estimated sampling variance (constructed from the averaged bootstrap distribution) to the true sampling variance of $\bar{y}_T$ in (12).

5.2 Simulation Results

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

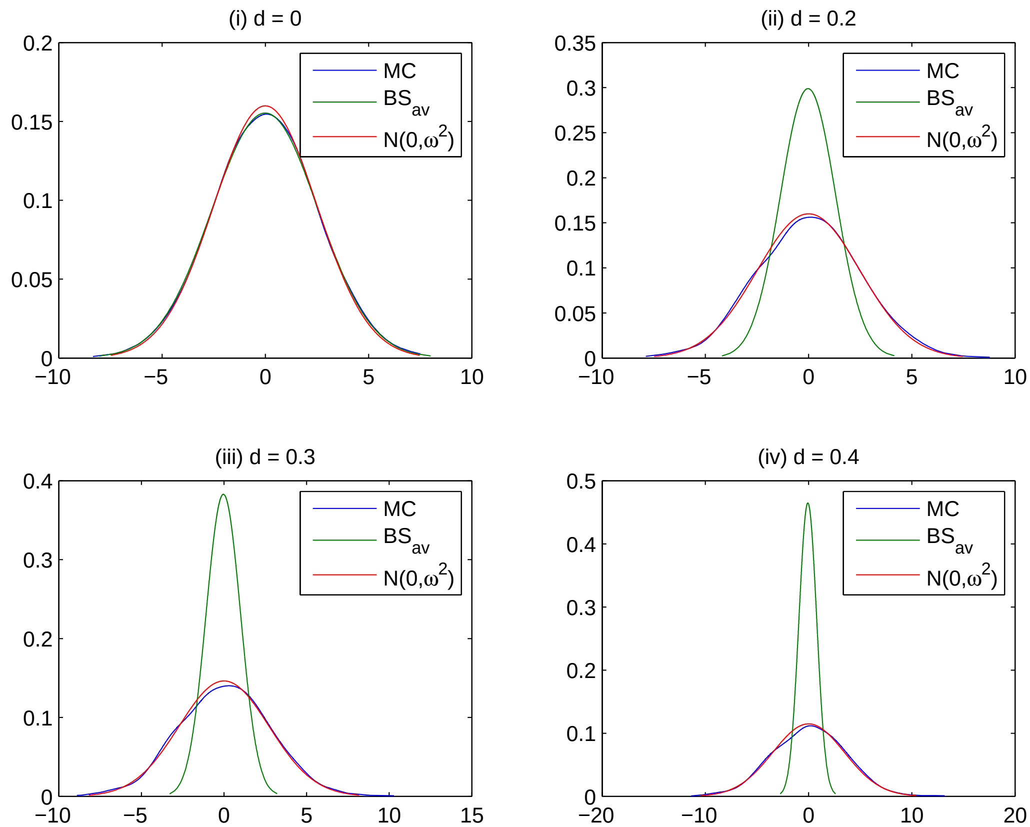

Figure 1: Density of the re-normalized sample mean for $T = 500$, $\phi = 0.6$, and $d = 0, 0.2, 0.3, 0.4$.

When $d = 0$ we can see that all three distributions are very nearly indentical. 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 bootstrap estimate of the standard deviation of $\bar{y}_T$ to the exact standard deviation given by the square root of (12), for both $T = 100$ and $T = 500$. For $d > 0$, all three values of ϕ and for both sample sizes, the underestimation is very marked with, indeed, the degree of underestimation increasing with ϕ and there being no uniform tendency for improvement as the sample size increases.[1]

[1]The mean and skewness of the re-nomalised 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. Andrews et al.

Table 1: Standard deviation of $\bar{y}_T$: averaged SB estimate as a percentage of $\sqrt{Var[\bar{y}_T]}$.

		d			
ϕ	T	0.0	0.2	0.3	0.4
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%
0.9	100	79.8%	58.5%	45%	30.1%
	500	96.3%	61.9%	46.8%	31.1%

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

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

where $\bar{\gamma}_h(k) = \widehat{\gamma}(k)$, $k = 0, 1, \ldots, h$, and $\sum_{j=0}^{h}\bar{\phi}_h(j)\bar{\gamma}_h(k-j) = 0$, $k = h+1, \ldots$, and Hosking (1996) shows that the $\widehat{\gamma}(k)$ 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 average value of $\widehat{\gamma}(k)$ for $k = 0, \ldots, 100$ obtained from samples of size $T = 1000$, computed from two fractional noise process 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

$$E[\widehat{\gamma}(k) - \gamma(k)] \sim -\omega^2 T^{2d-1}\,, \tag{14}$$

which depends on d but is independent of k, a feature that is reflected in the bias observed in the simulations, which is in close accord with the asymptotic approximation in (14), as can be seen in Figure 2.

Noting that $\frac{1}{T}\sum_{k=1-T}^{T-1}(1-\frac{|k|}{T}) \equiv 1$, it is apparent from equation (14) that the addition of $\omega^2 T^{2d-1}$ to the bootstrap variance would provide an asymptotically valid correction that would compensate for the bias of the $\widehat{\gamma}(k)$ and the underestimation of the true persistence

(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 $\bar{y}_T$. Our results indicate that approximating the sampling distribution of the sample mean is not straightforward, the symmetry and light tails of the normal distribution that may enhance the performance of the sieve bootstrap notwithstanding. 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.

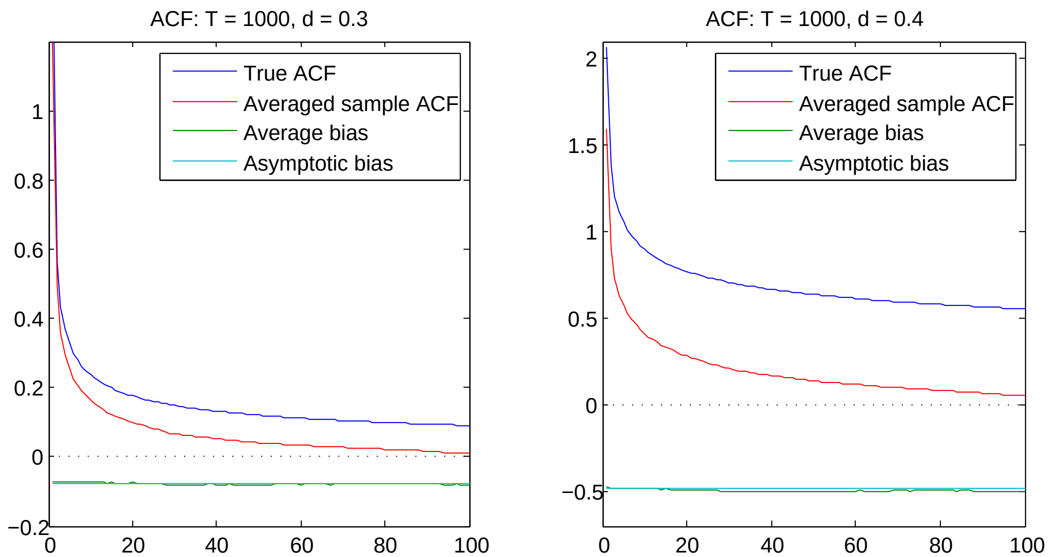

Figure 2: Theoretical and average sample autocovariance function, and asymptotic and average bias of fractional noise with $d = 0.3, 0.4$ and $T = 1000$.

in the observed process. We have therefore rescaled the raw SB draws of the re-normalized sample mean such that their variance is increased by ω^2. A subset of the averaged (over the MC draws) re-scaled BS distributions is displayed in Figure 3. We see that, with the exception of $d = 0.0$ (for which the variance inflation is invalid), this averaged re-scaled distribution (denoted by SBS hereafter) provides a much more accurate estimate of the true sampling distribution than does the (unscaled) raw SB. This conclusion, based on visual inspection of the relevant distributions, is confirmed by the numerical results in Table 2, which reports the averaged SBS estimate of $\sqrt{Var\left(\bar{y}_T\right)}$ as a percentage of its true value. Table 2 indicates that the ratio of the estimated to true standard error is reasonably close

Table 2: Standard deviation of $\bar{y}_T$: averaged rescaled SBS estimate as a percentage of $\sqrt{Var[\bar{y}_T]}$.

		d			
ϕ	T	0.0	0.2	0.3	0.4
0.3	100	138.6%	115.3%	108.7%	103.9%
	500	140.9%	111.2%	106.0%	102.6%
0.6	100	137.5%	116.9%	110.2%	104.7%
	500	140.9%	112.5%	106.6%	102.8%
0.9	100	131.9%	117.7%	110.8%	105.0%
	500	139.5%	117.8%	110.5%	104.8%

to 100% in all parts of the parameter space in which $d > 0$, even for the smaller sample size of $T = 100$.

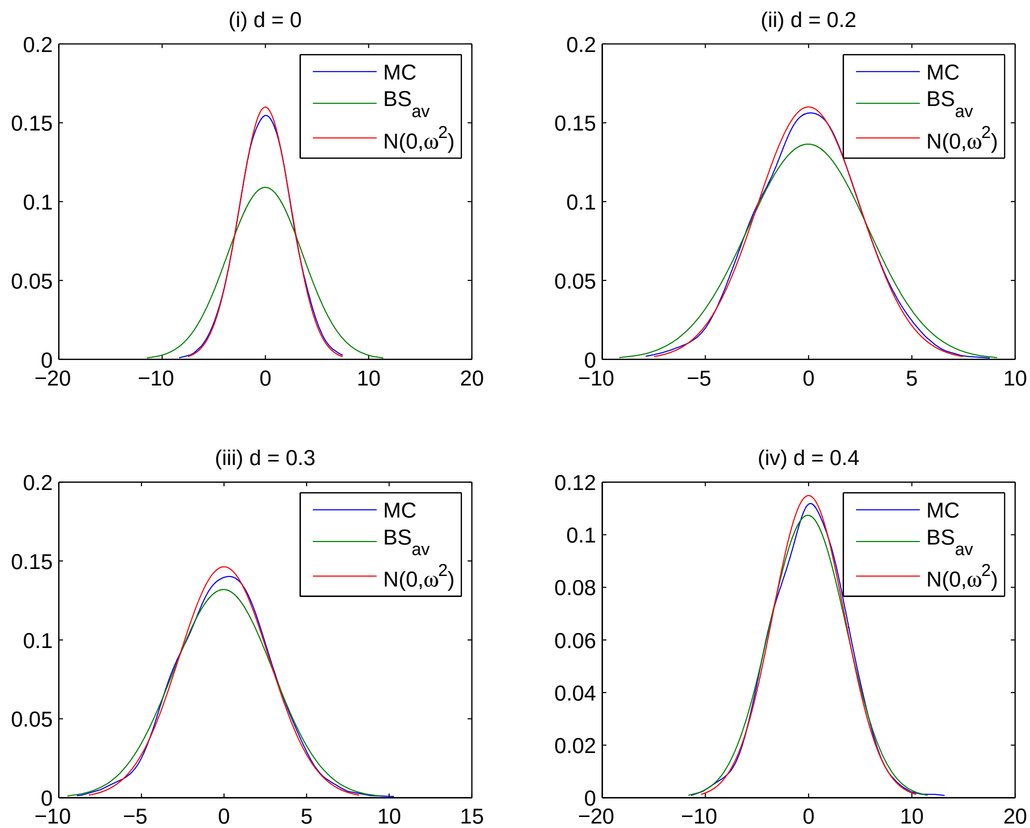

Figure 3: Density of the re-normalized sample mean for $T = 500$, $\phi = 0.6$, and $d = 0, 0.2, 0.3, 0.4$; with the bootstrap distributions rescaled before averaging.

We emphasize, however, that the SBS re-scaling - in addition to being strictly appropriate only for large sample sizes (and for $d > 0$) - is based on the use of the true values of d and $\kappa(1)$ in the computation of the scaling factor and is empirically infeasible as a consequence. A plug in estimate could be employed, with ω^2 replaced by

$$\widehat{\omega}^2 = \frac{\bar{\sigma}_h^2 \Gamma(1-2\widehat{d})}{\{\sum_{j=0}^{h} \bar{\phi}_h(j)\}^2 (1+2\widehat{d})\Gamma(1+\widehat{d})\Gamma(1-\widehat{d})} \,,$$

but preliminary investigations indicate that sampling error in the scaling factors renders this method ineffective. As such, it is of interest to ascertain whether the feasible PFSB algorithm, in *implicitly* producing more accurate estimates of the $\gamma(k)$ in the process of yielding bootstrap draws of $\bar{y}_T$, yields an estimated sampling distribution for the mean with a variance that is closer to the theoretical value, without any subsequent re-scaling of the draws.

Figure 4 graphs the Monte-Carlo distribution of $T^{1/2-d}(\bar{y}_T - \mu)$, the PFBS distribution of $T^{1/2-d}(\bar{y}_T^* - \bar{y}_T)$, and the $N(0, \omega^2)$ distribution for $T = 500$, $\phi = 0.6$, and $d = 0, 0.2, 0.3, 0.4$. As noted above, the PFSB algorithm is based on a (bias-adjusted) pre-filtering value of d

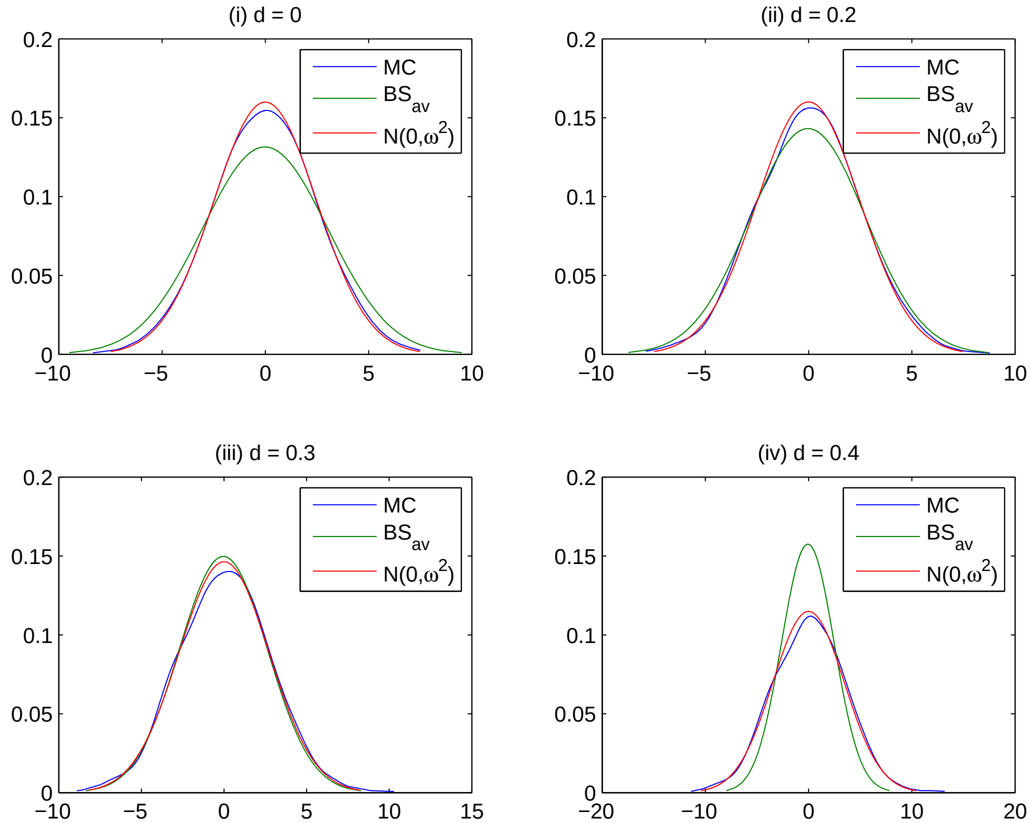

Figure 4: Density of the re-normalized sample mean for $T = 500$, $\phi = 0.6$, and $d = 0, 0.2, 0.3, 0.4$.

that is deemed to be 'optimal' in the matching experimental design in Poskitt et al. (2012). We see that overall, the PFSB algorithm provides a reasonable approximation to the true distribution, particularly vis-à-vis the variance of $\bar{y}_T$, and that apart from the $d = 0.4$ case the PFSB distributions are not too dissimilar from the SBS distributions. The failure of the PFSB to mimic the SBS when $d = 0.4$ is explicable in the fact, partially demonstrated in Figure 2, that ω^2 increases ever more rapidly as d increases from values below 0.25 to those above 0.25, so that the absolute bias in the fractional noise case when $T = 1000$, for example, is less than 0.015 when $d = 0.2$, about 0.075 when $d = 0.3$ and exceeds 0.45 when $d = 0.4$. Thus, whereas the infeasible SBS operates appropriately using the asymptotically correct bias adjustment, the PFSB has difficulty correcting for the large but unknown values of the bias that occur as d increases. Nevertheless, despite a tendency to slightly over-estimate $Var(\bar{y}_T)$ for small d (i.e., $d = 0, 0.2$), and under-estimate for large d ($d = 0.4$), the PFSB results are superior to those associated with the raw SB, and much closer (than are the raw SB results) to the (empirically unattainable) results based on re-scaling.

The averaged PFBS estimate of $\sqrt{Var\,(\bar{y}_T)}$ as a percentage of the true $\sqrt{Var\,(\bar{y}_T)}$ is presented in Table 3. The reasonable accuracy observed visually in Figure 4 for $\phi = 0.6$ is

Table 3: Standard deviation of $\bar{y}_T$: averaged PFBS estimate as a percentage of $\sqrt{Var[\bar{y}_T]}$.

		d			
ϕ	T	0.0	0.2	0.3	0.4
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%
0.9	100	401.8%	714.7%	941.5%	697.0%
	500	258.2%	227.0%	198.0%	150.7%

broadly replicated for $\phi = 0.3$, augering well for the automated use of this method in practice. We note that the very poor results for $\phi = 0.9$ reflect the fact that, like many other procedures, the PFBS algorithm does not perform well in this part of the parameter space, as is documented in Poskitt et al. (2012).

6 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 second moments. Prefiltering 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.

References

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 for the Whittle maximum likelihood estimator for stationary long-memory gaussian time series. *Econometric Theory* **21** 710–734.

Apostol, T. M. (1960). *Mathematical Analysis*. Addison-Wesley, Reading.

Beran, J. (1994). *Statistics for Long Memory Processes*. 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*. 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.

Grose, S. D. and Poskitt, D. S. (2006). The finite-sample properties of autoregressive approximations of fractionally-integrated and non-invertible processes. Working paper 15/06, Department of Econometrics and Business Statistics, Melbourne, Australia.

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.

Künsch, H. R. (1989). The jacknife 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.

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.

Sowell, F. (1992). Maximum likelihood estmation 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.