ISSN 1440-771X

MONASH University

Department of Econometrics and Business Statistics

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

Bias Reduction of Long Memory Parameter Estimators via the Pre-filtered Sieve Bootstrap

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

April 2012

Working Paper 08/12

Bias Reduction of Long Memory Parameter Estimators via the Pre-filtered Sieve Bootstrap

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

Department of Econometrics & Business Statistics, Monash University

Abstract

This paper investigates the use of bootstrap-based bias correction of semi–parametric estimators of the long memory parameter in fractionally integrated processes. The re-sampling method involves the application of the sieve bootstrap to data pre-filtered by a preliminary semi-parametric estimate of the long memory parameter. Theoretical justification for using the bootstrap techniques to bias adjust log periodogram and semi-parametric local Whittle estimators of the memory parameter is provided. Simulation evidence comparing the performance of the bootstrap bias correction with analytical bias correction techniques is also presented. The bootstrap method is shown to produce notable bias reductions, in particular when applied to an estimator for which analytical adjustments have already been used. The empirical coverage of confidence intervals based on the bias-adjusted estimators is very close to the nominal, for a reasonably large sample size, more so than for the comparable analytically adjusted estimators. The precision of inferences (as measured by interval length) is also greater when the bootstrap is used to bias correct rather than analytical adjustments.

Keywords: Analytical bias correction, bootstrap bias correction, confidence interval, coverage, precision, log periodogram estimator, local Whittle estimator.

JEL Classification: C18, C22, C52

1 Introduction

The so-called long memory, or strongly dependent, processes have come to play an important role in time series analysis. Long-range dependence, observed in a very wide range of empirical applications, is characterized by an autocovariance structure that decays too slowly to be absolutely summable. Specifically, rather than the autocovariance function declining at the exponential rate characteristic of a stable and invertible $ARMA$ process, it declines at

*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. The authors would like to acknowledge the constructive comments of participants at: the Symposium of Econometric Theory and Applications (Monash University, 2011); and the Australasian Econometric Society Meeting (University of Adelaide, 2011).

a hyperbolic rate dependent on a "long memory" parameter $H \in (0,1)$; i.e.,

$$\gamma(\tau) \sim C\tau^{-H}, C \neq 0, \text{ as } \tau \to \infty\ . \quad (1)$$

A detailed description of the properties of such processes can be found in Beran (1994).

Perhaps the most popular model of a long memory process is the fractionally integrated ($I(d)$) process introduced by Granger and Joyeux (1980) and Hosking (1980). This class of processes can be characterized by the specification,

$$y(t) = \sum_{j=0}^{\infty} k(j)\varepsilon(t-j) = \frac{\kappa(z)}{(1-z)^d}\,\varepsilon(t), \quad (2)$$

where $\varepsilon(t)$ is zero-mean white noise, z is here interpreted as the lag operator ($z^j y(t) = y(t-j)$), and $\kappa(z) = \sum_{j\geq 0} \kappa(j)z^j$, $\kappa(0) = 1$. For any $d > -1$ the operator $(1-z)^d$ is defined via the binomial expansion

$$(1-z)^d = 1 - dz + \frac{d(d-1)z}{2!} - \frac{d(d-1)(d-2)z^3}{3!} + \cdots\ , \quad (3)$$

and if the "short memory" component $\kappa(z)$ is the transfer function of a stable, invertible ARMA process and $|d| < 0.5$, then the coefficients of $k(z)$ are square–summable ($\sum_{j\geq 0} |k(j)|^2 < \infty$). In this case $y(t)$ is well-defined as the limit in mean square of a covariance-stationary process and the model is essentially a generalization of the classic Box-Jenkins ARIMA model (Box and Jenkins, 1970),

$$(1-z)^d \Phi(z) y(t) = \Theta(z)\varepsilon(t), \quad (4)$$

in which we now allow non-integer values of the integrating parameter d and $\kappa(z) = \Theta(z)/\Phi(z)$. The long run behaviour of this process naturally depends on the fractional integration parameter d. In particular, for any $d > 0$ the impulse response coefficients of the Wold representation in (2) are not absolutely summable and, for $0 < d < 0.5$, the autocovariances decline at the rate $\gamma(\tau) \sim C\tau^{2d-1}$ (i.e. with reference to (1), $H = 1-2d$). Such processes have been found to exhibit dynamic behaviour very similar to that observed in many empirical time series. See Robinson (2003) for a collection of the seminal articles in the area and Doukhan, Oppenheim and Taqqu (2003) for a thorough review of theory and applications. The role played by fractional processes in finance, most notably in the modelling of the variance of financial returns, is highlighted in Andersen, Bollerslev, Christoffersen and

Diebold (2006) and in multiple papers published in a special issue of *Econometric Reviews* (2008, 27, Issue 1-3).

Statistical procedures for analyzing long memory processes have ranged from the likelihood-based methods of Fox and Taqqu (1986), Dahlhaus (1989), Sowell (1992) and Beran (1995), to the semi-parametric techniques advanced by Geweke and Porter-Hudak (1983) and Robinson (1995a,b), among others. The asymptotic theory for maximum likelihood estimation (MLE) of the parameters of such processes is well established, at least under the assumption of Gaussian errors. In particular, we have consistency, asymptotic efficiency, and asymptotic normality for the MLE of the fractional differencing parameter, so providing a basis for large sample inference in the usual manner. Such asymptotic results are, however, conditional on correct model specification, with the MLE of d typically inconsistent if either or both the autoregressive and moving average operators in (4) (or, alternatively, the operator $\kappa(z)$ in (2)) are incorrectly specified. The semi-parametric methods aim to produce consistent estimators of d whilst placing only very mild restrictions on the behaviour of $\kappa(e^{\imath\lambda})$ for frequency values λ near zero. The semi-parametric estimators are therefore robust to different forms of short-run dynamics and offer broader applicability than a fully parametric method. They are also asymptotically pivotal and have particularly simple asymptotic normal distributions.

Whilst such features place the semi-parametric methods at the forefront for use in conducting inference on d, the price paid for their application is a reduction in asymptotic efficiency (relative to exact MLE) and a slower rate of convergence to the true parameter (Giraitis, Robinson and Samarov, 1997). Also, despite asymptotic robustness to the short-run dynamics, semi-parametric estimators have been shown to exhibit large finite sample bias in the presence (in particular) of a substantial autoregessive component – see Agiakloglou, Newbold and Wohar (1993) and Lieberman (2001) for example. Hence, bias-correction of semi-parametric estimators is an important area to explore.

In this paper we focus on bias-correction of the following two semi-parametric estimators $\widehat{d}_T$ of d:

1. The Geweke and Porter-Hudak (1983)/Robinson (1995b) log periodogram regression estimator (referred to hereafter as LPR): The ordinary least squares (OLS) slope coefficient in a regression of $\log I_T(\lambda_j)$ on a constant and $-2\log\lambda_j$, $j = 1, \ldots, N$,

$$\widehat{d}_T = \arg\min_{|d|<1/2} \sum_{j=1}^{N} \left(\log I_T(\lambda_j) + 2d(\log\lambda_j - \overline{\log\lambda})\right)^2 ,$$

where $I_T(\lambda) = (2\pi T)^{-1}|\sum_{t=1}^{T} y(t)e^{-\imath\lambda}|^2$, the periodogram, $\lambda_j = 2\pi j/T$, $j = 1, \ldots, N$, are the first N fundamental frequencies, and $\overline{\log\lambda} = N^{-1}\sum_{j=1}^{N}\log\lambda_j$.[1]

2. The semi-parametric Gaussian (local Whittle) estimator of Robinson (1995a) (SPLW):

$$\widehat{d}_T = \arg\min_{|d|<1/2}\left(\log(N^{-1}\sum_{j=1}^{N}\lambda_j^{2d} I_T(\lambda_j)) - 2d\overline{\log\lambda}\right).$$

Both $\widehat{d}_T$ are $\sqrt{N}$–CAN estimators of d, by which we mean that $\widehat{d}_T$ is consistent, $|\widehat{d}_T - d| = o(1)$, and asymptotically normal, $|P\left(N^{1/2}(\widehat{d}_T - d)/\upsilon < x\right) - G(x)| = o(N^{-1/2})$ where $G(x)$ denotes the standard normal cumulative distribution function. For the LPR estimator $\upsilon^2 = \pi/24$, and $\upsilon^2 = 1/4$ for the SPLW estimator. For both estimators the bandwidth parameter N, denoting the number of periodogram ordinates employed, is chosen as a monotonically increasing function of sample size T , and because $\kappa(z)$ is only specified locally, N must be assigned such that $N/T \to 0$ as $T \to \infty$. Too small a choice of N may prompt concern about the accuracy of the normal approximation, whereas too large a value for N entails an element of non–local averaging and is a source of bias. In brief, although $\lim_{T\to\infty} E[\widehat{d}_T - d] = 0$ the finite sample bias in such estimators can, as previously observed, present problems.

One approach to the problem of bias is to seek an analytical solution that will reduce the first-order bias. Andrews and Guggenberger (2003), for example, consider a bias-adjusted estimator of d obtained by including even powers of frequency as additional regressors in the log-periodogram pseudo regression, and Andrews and Sun (2004) adapt this approach to the SPLW estimator. Moulines and Soulier (1999) reduce bias by constructing a broad-band estimator of d that uses all of the frequencies in the range $(0, \pi]$, not just those in a neighborhood of zero. Monte–Carlo evidence presented in Nielsen and Frederiksen (2005) demonstrates the usefulness of the bias-adjusted LPR and SPLW estimators. In particular, the bias-corrected semi-parametric estimators are shown to outperform correctly specified parametric estimators, although at the expense of an increase in mean squared error.

An alternative methodological approach to bias-correction, and the one that we examine here, is to use the bootstrap. Bootstrap methodology may be thought of as coming in two "flavours" – the parametric, or model–based, bootstrap, and a variety of non– or semi–parametric schemes. The parametric bootstrap relies on having a full, correct parametric specification for the process and is therefore at odds with the semi–parametric approach to estimation being considered here. A less model-dependent approach nonetheless requires a

[1] We have written the estimator in the form given by Robinson (1995b). Geweke and Porter-Hudak (1983) use the regressor $-2\log|1 - e^{-\imath\lambda_j}|$. The two are equivalent because $|1 - e^{-\imath\lambda}|^{2d} = |\lambda|^{2d}(1 + o(1))$ as $\lambda \to 0$.

re-sampling scheme that is able to capture the salient features of the data generating process, the dependence structure of the process being of prime importance in the time series context. While the block bootstrap of Künsch (1989) has traditionally been employed for this purpose, blocking has been found to suffer from relatively poor convergence rates. For instance, the error in the coverage probability of a one-sided confidence interval derived from the block bootstrap is $O(T^{-3/4})$, compared to the $O(T^{-1})$ rate achieved with simple random samples. An attractive alternative is the "sieve" bootstrap of Bühlmann (1997). This works by "pre–whitening"the data using an autoregressive approximation, with the dynamics of the process captured in a fitted autoregression. Provided the order, h, of the autoregression increases at a suitable rate with T, the convergence rates for the sieve bootstrap are much closer (in fact arbitrarily close) to those for simple random samples. Choi and Hall (2000) demonstrate the superior convergence performance of the sieve bootstrap (over the block bootstrap) for linear short memory processes, whilst Poskitt, Grose and Martin (2012) build on the results of Poskitt (2008) to show that under regularity conditions that allow for fractionally integrated $I(d)$ processes the sieve bootstrap achieves an error rate of $o(T^{-(1-\max\{0,d\})+\beta})$ for all $\beta > 0$ for a general class of statistics.

The current paper uses a modified sieve bootstrap to bias-correct the LPR and SPLW estimators of the memory parameter in fractionally integrated $I(d)$ processes. The bootstrap method uses a consistent semi-parametric estimator of the long memory parameter to pre-filter the raw data, *prior to* the use of a long autoregressive approximation as the 'sieve' from which bootstrap samples are produced. The bias correction proceeds in an iterative fashion, with a stochastic stopping rule invoked to produce the final estimator. Starting with the Edgeworth expansions of Giraitis and Robinson (2003), and using the convergence properties determined in Poskitt et al. (2012), we derive error rates for estimating the bias of both the LPR and SPLW estimators, with the accuracy with which the bootstrap method estimates the bias in finite samples then documented in a simulation setting. We explore the impact of using two different semi-parametric estimates of d in the pre-filtering, including analytical bias adjustments thereto. In summary, we document the sampling properties of the following estimators:

1. The pre-filtered sieve bootstrap bias-corrected LPR estimator, based on the LPR pre-filter (LPR_{sb})

2. The pre–filtered sieve bootstrap bias corrected SPLW estimator, based on the SPLW pre-filter, (SPLW_{sb})

3. The pre-filtered sieve bootstrap bias-corrected LPR estimator, based on the analytically bias-corrected LPR pre-filter (LPR-BA$_{sb}$)

4. The pre–filtered sieve bootstrap bias corrected SPLW estimator, based on the analytically bias-corrected SPLW pre-filter (SPLW-BA$_{sb}$)

The analytically bias-adjusted LPR estimator of Andrews and Guggenberger (2003) (LPR-BA) is produced as the OLS coefficient of the regressor $-2\log\lambda_j$ in the regression of $\log I_T(\lambda_j)$ on $\overline{\log\lambda}$, $-2\log\lambda_j$, and λ_j^{2p}, $p = 1,\ldots,P$, $j = 1,\ldots,N$. The analytically bias-adjusted SPLW estimator of Andrews and Sun (2004) (SPLW-BA) is produced as the first element of $(\widehat{d}_T, \widehat{\theta}_1, \ldots, \widehat{\theta}_P) = \arg\min LW(d, \theta_1, \ldots, \theta_P)$ where

$$LW(d,\theta_1,\ldots,\theta_P) = \log\left(N^{-1}\sum_{j=1}^{N}\lambda_j^{2d}I_T(\lambda_j)\exp\left\{\sum_{p=1}^{P}\theta_p\lambda_j^{2p}\right\}\right) - N^{-1}\sum_{j=1}^{N}\left\{\sum_{p=1}^{P}\theta_p\lambda_j^{2p}\right\} - 2d\overline{\log\lambda}\,.$$

The accuracy of the bootstrap bias-corrected estimators in 1 and 2 is assessed against that of the corresponding analytically bias-adjusted estimators in simulation experiment. In brief, using a mean squared error criterion, the iterative bootstrap method is shown to produce more finite sample accuracy than the analytically adjusted counterparts. Furthermore, with the analytically bias-adjusted estimators used as pre-filters (in 3 and 4), further finite sample accuracy is gained via the bootstrap method.

The paper proceeds as follows. Section 2 briefly summarizes the statistical properties of long memory processes, and outlines the sieve bootstrap (both 'raw' and pre-filtered) in this context. The pre–filtered sieve bootstrap bias adjustment (PFBS(BA)) algorithm is also described in this section. In Section 3 we present the relevant approximations and exploit these to produce the error rates for the bootstrap technique. Section 4 outlines the iterated version of the bootstrap bias correction technique. Details of the simulation study are given in Section 5. Section 6 closes the paper.

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 representation as in (2) where;

Assumption 1. *The transfer function in the Wold representation is given by* $k(z) = \kappa(z)/(1-z)^d$ *with* $|z| \leq 1$, $|d| < 0.5$ *and* $\kappa(z) \neq 0$. *The impulse response coefficients of* $\kappa(z)$ *satisfy* $k(0) = 1$ *and* $\sum_{j \geq 0} j|\kappa(j)| < \infty$.

Assumption 2. *The innovations process* $\varepsilon(t)$ *is ergodic and,*

$$E\left[\varepsilon(t) \mid \mathcal{E}_{t-1}\right] = 0 \quad \textit{and} \quad E\left[\varepsilon(t)^2 \mid \mathcal{E}_{t-1}\right] = \sigma^2\,, \tag{5}$$

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

Assumption 1 incorporates quite a wide class of linear processes, including the ARFIMA family of models that are the focus of this paper. Assumption 2 imposes a classical martingale difference structure on the innovations; the fundamental 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. Weaker mixing conditions could be employed and Assumption 1 could also be relaxed, but we will not investigate such generality here.

Under Assumption 2 the linear predictor

$$\bar{y}(t) = \sum_{j=1}^{\infty} \pi(j) y(t-j)\,, \tag{6}$$

where $\sum_{j=0}^{\infty} \pi(j) z^j = (1-z)^d \kappa(z)^{-1}$, is the minimum mean squared error (MMSE) predictor (MMSEP) of $y(t)$. The MMSEP of $y(t)$ based only on a finite number (h) of past observations (MMSEP(h)) is then

$$\bar{y}_h(t) = \sum_{j=1}^{h} \pi_h(j) y(t-j) \equiv -\sum_{j=1}^{h} \phi_h(j) y(t-j); \tag{7}$$

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

$$\varepsilon_h(t) = \sum_{j=0}^{h} \phi_h(j) y(t-j). \tag{8}$$

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

$$\sum_{j=0}^{h} \phi_h(j)\gamma(j-k) = \delta_0(k)\sigma_h^2\,, \quad k = 0, 1, \ldots, h, \tag{9}$$

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 the MMSE is

$$\sigma_h^2 = E\left[\varepsilon_h(t)^2\right], \tag{10}$$

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 that 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 conventional $I(0)$ 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, with order selected via an asymptotically efficient criterion, is accordingly a plausible "non-parametric" bootstrap technique for long-memory processes.

2.1 The raw sieve bootstrap

Details of the raw sieve bootstrap (SB) for fractional processes are given in Poskitt (2008). For convenience we reproduce here the basic steps of the SB algorithm for generating a realization of a process $y(t)$, prior to presenting the PFBS(BA) algorithm adopted for bias-adjustment in this paper.

The raw sieve bootstrap (SB) algorithm:

SB1. Given data $y(t)$, $t = 1, \ldots, T$, and using $y(1-j) = y(T-j+1)$, $j = 1, \ldots, h$, as initial values, 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

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

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

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

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 is a discrete uniform random variable with support on the integers $h, \ldots, T$.

Crucially, 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, as formalized in the following theorem

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 linearly regular, covariance-stationary process that satisfies Assumptions 1 and 2, then for all $h \le H_T = a(\log T)^c$, $a > 0$, $c < \infty$,*

$$\sum_{j=1}^{h} |\bar{\phi}_h(j) - \phi_h(j)|^2 = O\left(h\left(\frac{\log T}{T}\right)^{1-2\,max\{0,d\}}\right).$$

The proof of this theorem is placed in the Appendix of Proofs, along with the proofs of other results presented in the paper.

2.2 The pre-filtered sieve bootstrap

Given the dependence of the convergence of $\sum_{j=0}^{h} \bar{\phi}_h(j)z^j$ to $\sum_{j=0}^{h} \phi_h(j)z^j$ on the value of d, the convergence of any bootstrap generated sampling distribution to the true unknown sampling distribution is also dependent on the value of d, see Poskitt (2008). In particular, in Poskitt et al. (2012) it is shown that under appropriate regularity the raw sieve bootstrap achieves a convergence rate of $o(T^{-(1-\max\{0,d\})+\beta})$ for all $\beta > 0$. Obviously, in the long memory case where $0 < d < 0.5$, the closer is d to zero the closer the convergence rate of $o(T^{-(1-d)+\beta})$ will be to the rate $o(T^{-1+\beta})$ achieved with short memory (and anti-persistent) processes. The empirical regularity of estimated values of d in the $0 < d < 0.5$ range thus provides motivation for the idea of pre-filtering the series prior to the application of the sieve. Specifically, we employ a modified sieve method wherein, for a given preliminary value of d, we pre-filter the data using this value, apply the AR approximation (and sieve bootstrap) to the pre-filtered data, before using the inverse filter to produce the final realization of $y(t)$. With this procedure, the raw sieve is applied (by construction) to filtered data with shorter memory; hence the achievement of an improved convergence rate. Moreover, Hosking (1996) shows that the sample autocorrelations of a long memory process have substantial negative bias relative to the corresponding true autocorrelations, even for moderate to large samples. Theoretical properties notwithstanding, this may also impinge on the performance of the sieve bootstrap, effectively based as it is on the estimation of sample autocorrelations. This provides further motivation for reducing the memory in the "data" to which the sieve is applied, via the pre–filtering procedure.

For any $d > -1$ let $(1-z)^d = \sum_{j=0}^{\infty} \alpha_j^{(d)} z^j$ where $\alpha_j^{(d)}$, $j = 0, 1, 2, \ldots$, denote the coefficients of the fractional difference operator when expressed in terms of its binomial expansion, as in the right hand side of 3. Given a preliminary value d^f of d, pre–filtered sieve bootstrap (PFSB) realizations of $y(t)$ are generated using the following algorithm:

PFBS1. Calculate the coefficients of the filter $(1-z)^{d^f}$ and from the data generate the filtered values

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

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

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

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

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

of $y(t)$ for $t = 1, \ldots, T$, where the superscript f is used to distinguish this bootstrap draw from the bootstrap draw produced by the raw sieve algorithm, in Step SB3 above.

In Poskitt et al. (2012) it is shown that given a judicious choice of d^f, shorter memory will be induced by the preliminary filtering at Step PFSB1. The accuracy of the AR approximation and, thereby, the sieve bootstrap in Step PFSB2 will accordingly be increased, and this increase in accuracy will be passed on to the PFSB draws in Step PFSB3, resulting in a convergence rate equal to that obtained in the short memory case, namely $o(T^{-1+\beta})$. With these results in hand, we proceed to work with the PFSB algorithm for the purpose of bias adjustment.

2.3 Bias correction via the pre-filtered sieve bootstrap

To bias adjust a chosen estimator, $\widehat{d}_T$, of d we proceed as follows:

BA1. Calculate $\widehat{d}_T$ from the data $y(t)$, $t = 1, \ldots, T$.

BA2. Use d^f as the preliminary value in Steps PFSB1-PSFB3 of the PFSB algorithm and produce B bootstrap realizations $y_b^{*f}(t)$, $t = 1, \ldots, T$, $b = 1, \ldots, B$, of the process $y(t)$. From these construct B bootstrap values of the estimator, $\widehat{d}_{T,b}^{*f}$, $b = 1, 2, ..., B$, by evaluating the estimator $\widehat{d}_T$ for each of the B independent bootstrap draws.

BA3. Estimate the bias of $\widehat{d}_T$ by

$$\widehat{b}_{T,B}^{*f} = \left(\frac{1}{B} \sum_{b=1}^{B} \widehat{d}_{T,b}^{*f} \right) - d^f \tag{11}$$

and produce the bias adjusted estimator

$$\widetilde{d}_T = \widehat{d}_T - \widehat{b}_{T,B}^{*f}. \tag{12}$$

We refer to this as the PFBS(BA) algorithm.

3 Some Theoretical Underpinnings

The use of the PFBS(BA) algorithm to correct the finite sample bias of an estimator is justified only if the method produces a bootstrap distribution that copies the true sampling distribution of the estimator to the appropriate order of magnitude. Not surprisingly, the rate of convergence of the bootstrap to the true sampling distribution is shown to be dependent on the proximity of the preliminary value employed in the PFBS, namely d^f, to the true value of d, as well as the order (h) of the autoregressive approximation used in the sieve component of the PFBS. Presuming that d^f is itself estimated from the data, $d^f = d_T^f$ say, the main content of these findings, for general $\sqrt{N}$–CAN estimators d_T^f and $\widehat{d}_T$, is presented in Theorems 7 and 8. However, we begin the section with more specific results (that provide some insight into the general results that follow) for the case of the SPLW estimator and any pre-filtering value d^f such that $|d^f| < 1/2$, $d^f - d \in N_\delta = \{x : |x| \leq \delta\}$ where $2\delta < \min\{|d|, 0.5 - |d|\}$.

Let $\widehat{d}_T$ denote the value of the SPLW estimator when calculated from a sample of size T using a bandwidth parameter $N \sim KT^{4/5}$ where $K \in (0, \infty)$. Now suppose, for the sake of argument, that $\varepsilon(t)$ is an i.i.d $N(0, \sigma^2)$ white noise process. Then the process $y(t)$ is Gaussian, $|d| < 1/2$ by assumption, and $|\kappa(e^{\imath\lambda})|^2 = c_0 + c_1\lambda^2 + c_2\lambda^4 + o(|\lambda|^4)$ as $\lambda \to 0$, remembering that $|\kappa(e^{\imath\lambda})|^2$ is a bounded, even function of λ by Assumption 1. It follows that as $T \to \infty$ the sampling distribution of $\widehat{d}_T$ admits the Edgeworth expansion

$$\sup_x |P\{2N^{1/2}(\widehat{d}_T - d + \beta K^{5/2}) < x\} - G(x)| = o(N^{-1/2}\log^4 N) \tag{13}$$

where

$$\beta = \frac{c_1}{N^{1/2} 36\pi^2 c_0}$$

(Giraitis and Robinson, 2003, Theorem 3.1).

Now let $\widehat{d}_T^{*f}$ denote the value of $\widehat{d}_T$ calculated from a bootstrap realization of the process, $y^{*f}(t)$, $t = 1, \ldots, T$, constructed using the PFSB algorithm where, once again for the sake of argument; (i) the preliminary value d^f satisfies the conditions stated above and (ii) the innovations $\varepsilon_h^*(t)$, $t = 1, \ldots, T$, used in Step PFSB2 are explicitly generated as i.i.d. $N(0, \bar{\sigma}_h^2)$. Since the process $\varepsilon_h^*(t)$ is now Gaussian, $y^{*f}(t)$ will also be Gaussian.[2] Set $\bar{\kappa}_h(z) = \sum_{j=0}^{\infty} \bar{\kappa}_h(j) z^j$

[2]The innovations generated in Step PFBS2 are i.i.d. $(0, \bar{\sigma}_h^2)$ by construction (See Steps SB1–SB2 of the SB algorithm). When the process $y(t)$ is Gaussian, $\varepsilon_h^*(t)$, $t = 1, \ldots, T$, and therefore $y^{*f}(t)$, $t = 1, \ldots, T$, will also be approximately Gaussian.

where the $\bar{\kappa}_h(j)$ and $\bar{\phi}_h(j)$ are related by the recursions

$$\bar{\phi}_h(0) = \bar{\kappa}_h(0) = 1\,, \quad \sum_{i=0}^{j} \bar{\kappa}_h(i)\bar{\phi}_h(j-i) = 0,\; j = 1, 2, \ldots\,. \tag{14}$$

By construction $\bar{\kappa}_h(z)\bar{\phi}_h(z) = 1$ for all $|z| \leq 1$ and $\bar{\kappa}_h(z)$ yields the AR(h) estimate of $\kappa(z)$ implicit in the PFSB and $|\bar{\kappa}_h(e^{\imath\lambda})|^2 = |\sum_{j=0}^{h} \bar{\phi}_h(j)e^{\imath\lambda j}|^{-2} = \bar{c}_0 + \bar{c}_1\lambda^2 + \bar{c}_2\lambda^4 + o(|\lambda|^4)$ as $\lambda \to 0$. It follows from Giraitis and Robinson (2003, Theorem 3.1) once more that as $T \to \infty$

$$\sup_x |P^*\{2N^{1/2}(\widehat{d}_T^{*f} - d^f + \bar{\beta}K^{5/2}) < x\} - G(x)| = o(N^{-1/2}\log^4 N), \tag{15}$$

where $\bar{\beta} = \bar{c}_1/(N^{1/2}36\pi^2\bar{c}_0)$.

Applying the triangular inequality, it follows from equations (13) and (15) that

$$\begin{aligned}\sup_x |P^*\{2N^{1/2}(\widehat{d}_T^{*f} - d^f) < x\} - P\{2N^{1/2}(\widehat{d}_T - d) < x\}| \leq \\ \sup_x |G(x + 2N^{1/2}\beta K^{5/2}) - G(x + 2N^{1/2}\bar{\beta}K^{5/2})| + o(N^{-1/2}\log^4 N)\,. \end{aligned}\tag{16}$$

But from the first mean value theorem for integrals (Apostol, 1960, Theorem 7.30)

$$\sup_x |G(x + 2N^{1/2}\beta K^{5/2}) - G(x + 2N^{1/2}\bar{\beta}K^{5/2})| \leq \frac{2N^{1/2}K^{5/2}}{\sqrt{2\pi}}|\beta - \bar{\beta}|$$

and we see that the difference between the PFSB distribution of $N^{1/2}(\widehat{d}_T^{*f} - d^f)$ and the sampling distribution of $N^{1/2}(\widehat{d}_T - d)$ is $O(N^{1/2}|\bar{\beta} - \beta|) + o(N^{-1/2}\log^4 N)$. The following lemma is used to specify $O(N^{1/2}|\bar{\beta} - \beta|)$ and, hence, the elements that determine the rate of convergence of the PFSB to the true sampling distribution of $N^{1/2}(\widehat{d}_T - d)$, where, once again, at this point $\widehat{d}_T$ is the SPLW estimator of d.

Lemma 4. *Suppose that the process $y(t)$ satisfies Assumptions 1 and 2, and that the PFSB algorithm is applied using a preliminary value d^f such that $d^f - d \in N_{\delta_T}$ where $\delta_T < \delta$ and $2\delta < \min\{|d|, 0.5 - |d|\}$, and an AR($h$) approximation. Then for all $h \leq H_T = a(\log T)^c$, $a > 0$, $c < \infty$,*

$$\lim_{T\to\infty} \left| |\bar{\kappa}_h(e^{\imath\lambda})|^2 - |\kappa(e^{\imath\lambda})|^2 \right| \leq \nu_{1,T} + \nu_{2,T} + \nu_{3,T}$$

where $\nu_{1,T} = \exp(\delta_T(\log T - \log\log T))O(h\,(\log T/T)^{1/2})$, $\nu_{2,T} = O(\delta_T h^{-|d|})$ and $\nu_{3,T} = O(\delta_T \log T)$, uniformly in λ for all $\lambda \in [2\pi/T, 2\pi N/T]$.

Lemma 4 indicates that within a small neighbourhood of the origin the order of magnitude of$|\bar{\kappa}_h(e^{\imath\lambda})|^2 - |\kappa(e^{\imath\lambda})|^2$ is $O(\nu_{1,T} + \nu_{2,T} + \nu_{3,T})$, or smaller, uniformly in λ. Simple algebraic manipulation also gives us the following bound

$$
\begin{aligned}
N^{1/2}|\bar{\beta}-\beta| &= \frac{1}{36\pi^2}\left|\frac{\bar{c}_1}{\bar{c}_0}-\frac{c_1}{c_0}\right| \\
&\leq \frac{1}{36\pi^2}\left(\left|\frac{c_1(\bar{c}_0-c_0)}{c_0\bar{c}_0}\right|+\left|\frac{(\bar{c}_1-c_1)}{\bar{c}_0}\right|\right),
\end{aligned}
$$

implying that the order of magnitude of $N^{1/2}|\bar{\beta}-\beta|$ is determined by that of $(\bar{c}_0-c_0)$ and $(\bar{c}_1-c_1)$. Writing $|\bar{\kappa}_h(e^{\imath\lambda})|^2-|\kappa(e^{\imath\lambda})|^2=(\bar{c}_0-c_0)+(\bar{c}_1-c_1)\lambda^2+(\bar{c}_2-c_2)\lambda^4+o(|\lambda|^4)$ and employing Lemma 4 hence leads to the following corollary.

Corollary 5. *Under the same conditions as for Lemma 4*

$$N^{1/2}|\bar{\beta}-\beta|=O(\nu_{1,T}+\nu_{2,T}+\nu_{3,T})\,.$$

Corollary 5 indicates that for the SPLW estimator calculated with a bandwidth $N\sim KT^{4/5}$ the convergence of $\bar{\beta}$ to β depends (as pre-empted above) on the order of the autoregressive approximation and the proximity of the preliminary value employed in the PFSB to the true d; namely the value of h and the value of δ_T implicit in the choice of d^f.

An optimal value of h can be achieved by selecting h using AIC, or an equivalent criterion. Denoting the said estimate by $\widehat{h}_{AIC}$, Poskitt (2007, Theorem 7) shows that in the fractional case – as in the short memory case – the figure of merit proposed by Shibata (1980), the integrated relative mean squared error

$$IRMSE(h)=\frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\frac{|\widehat{\kappa}_h(e^{\imath\lambda})|^2-|\widehat{\kappa}(e^{\imath\lambda})|^2}{|\widehat{\kappa}(e^{\imath\lambda})|^2}\right|^2,$$

is minimized asymptotically at $h=\widehat{h}_{AIC}$. This seems particularly pertinent in light of the source of the second remainder term that appears in Lemma 5. The use of $\widehat{h}_{AIC}$ yields an autoregressive order $h\sim\log T$.

Appropriate selection of the preliminary value for d is less clear. From Lemma 5 we can see that we require d^f to be such that $|d^f-d|\log T=o(1)$, but no other features of the result nor its derivation give us a guide as to the most appropriate choice of d^f. The argument preceding Lemma 5 is based on the work of Giraitis and Robinson (2003), which assumes Gaussianity, but Lemma 5 itself only relies on Assumptions 1 and 2, suggesting that similar results might be applicable more generally. Also, although the value of $N\sim KT^{4/5}$ assumed by Giraitis and Robinson (2003) gives the optimal MSE value, asymptotic normality of the SPLW estimator (and other estimators considered here) requires that $N=o(T^{4/5})$ (See the discussion in

Giraitis and Robinson, 2003, pp. 1333–1335). Henceforth we will, therefore, remove both the assumption that the data generating processes are Gaussian and the assumption that the bandwidth is specified via the optimal MSE value. In the developments that follow, we also characterize the difference between the bootstrap and true sampling distributions directly in terms of finite sample bias.[3]

In order to generalize the previous argument suppose that the pre-filtering value $d^f = d^f_T$, is a $\sqrt{N}$–CAN estimator of d. For any ϵ, $0 < \epsilon < 0.5$, it follows from the tail area properties of the normal distribution that $\lim_{T\to\infty} P(|d^f_T - d| > N^{-1/2+\epsilon}) \leq \exp(-N^{2\epsilon}/2\upsilon)$. 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 $|d^f_T - d| = o(N^{-1/2+\epsilon})$ and hence that $|d^f_T - d|\log T = o(N^{-1/2+\epsilon}\log T)$. It follows that $|d^f_T - d|\log T$ will be $o(1)$, as required, if $\log T/N^{1/2-\epsilon} \to 0$ as $T \to \infty$, for the chosen bandwidth N.

Now assume that $\widehat{d}_T$ (the estimator to be bias-corrected) is also a $\sqrt{N}$–CAN estimator that satisfies the requirement that $N^{-1/2+\epsilon}\log T = o(1)$, and let b_T denote the finite sample bias of $\widehat{d}_T$. Then $E[\widehat{d}_T] = d + b_T$, and since $\lim_{T\to\infty} N^{1/2}E[\widehat{d}_T - d] = 0$ we have $N^{1/2}b_T \to 0$ and $b_T \log N \to 0$. Here E denotes expectation taken with respect to the original probability space $(\Omega, \mathfrak{F}, P)$.

Theorem 6. *(Chibisov, 1972, Theorem 2) If V_T and U_T are two random variables such that $V_T = U_T + \xi_T$ where $P(|\xi_T| > \rho_N\sqrt{N}) = o(N^{-1/2})$ for some sequence $\rho_N \to 0$ then the Edgeworth expansion of V_T is equal to that of U_T up to order $o(N^{-1/2})$*

Substituting $V_T = \widehat{d}_T - d$, $U_T = \widehat{d}_T - E[\widehat{d}_T]$, $\xi_T = b_T$ and $\rho_N = N^{-1/2}\log N$ into Theorem 6 it follows that

$$\sup_x \left| P\left(\frac{N^{1/2}(\widehat{d}_T - E[\widehat{d}_T])}{\upsilon} < x\right) - P\left(\frac{N^{1/2}(\widehat{d}_T - d)}{\upsilon} < x\right)\right| = o(N^{-1/2})\,. \tag{17}$$

Similarly, if $(\Omega^*, \mathfrak{F}^*, P^*)$ denotes the probability space induced by the bootstrap process and E^* the associated expectation, replacing $\widehat{d}_T$ by $\widehat{d}^{*f}_T$, d by d^f_T and $E[\widehat{d}_T]$ by $E^*[\widehat{d}^{*f}_T] = d^f_T + \bar{b}_T$ (with $\bar{b}_T = E^*[\widehat{d}^{*f}_T] - d^f_T$ by construction)**addition here**, we also obtain

$$\sup_x \left| P^*\left(\frac{N^{1/2}(\widehat{d}^{*f}_T - E^*[\widehat{d}^{*f}_T]}{\upsilon} < x\right) - P^*\left(\frac{N^{1/2}(\widehat{d}^{*f}_T - d^f_T)}{\upsilon} < x\right)\right| = o(N^{-1/2}) \tag{18}$$

via a repetition of the above argument using Theorem 6.

[3]Note that larger bandwidth, as implied by the use of a value of $N \sim KT^{4/5}$, entails larger bias.

Theorem 7. *Suppose that the process $y(t)$ satisfies Assumptions 1 and 2, and that the PFSB algorithm is applied to $\widehat{d}_T$ using the preliminary value d_T^f and an AR(h) approximation. Assume that d_T^f and $\widehat{d}_T$ are $\sqrt{N}$–CAN estimators with bandwidth parameter chosen such that $N^{-1/2+\epsilon}\log T = o(1)$, $0 < \epsilon < 0.5$. Then for all $h \leq H_T = a(\log T)^c$, $a > 0$, $c < \infty$,*

$$\sup_x |P\{N^{1/2}(\widehat{d}_T - E[\widehat{d}_T]) < x\} - P^*\{N^{1/2}(\widehat{d}_T^{*f} - E^*[\widehat{d}_T^{*f}]) < x\}| = O(N^{1/2}|b_T - \bar{b}_T|) + o(N^{-1/2})\,.$$

Theorem 7 indicates that if d_T^f and $\widehat{d}_T$ are $\sqrt{N}$–CAN estimators with an appropriately chosen bandwidth, then the difference between the bootstrap distribution and the true sampling distribution depends on the difference between the true finite sample bias, b_T, and the estimate of this bias from the PFSB distribution, $\bar{b}_T$. This result parallels that already observed with the SPLW estimator. The following theorem indicates that the $N^{1/2}\left|\bar{b}_T - b_T\right|$ term plays the same role as the corresponding term $N^{1/2}\left|\bar{\beta} - \beta\right|$ plays in the SPLW case; see equation (16).

Theorem 8. *Suppose that the conditions in Theorem 7 hold. Assume that there exists a function $\beta_T(\cdot)$ such that $b_T = E[\widehat{d}_T] - d = N^{-1/2}\beta_T(c_0,\ldots,c_P,d) + o(N^{-1/2})$ where $|\kappa(e^{\imath\lambda})|^2 = \sum_{p=0}^{P} c_p\lambda^{2p} + o(|\lambda|^{2P})$ as $\lambda \to 0$ and $\bar{b}_T = E[\widehat{d}_T^{*f}] - d_T^f = N^{-1/2}\beta_T(\bar{c}_0,\ldots,\bar{c}_P,d_T^f) + o(N^{-1/2})$ where $|\bar{\kappa}(e^{\imath\lambda})|^2 = \sum_{p=0}^{P}\bar{c}_p\lambda^{2p} + o(|\lambda|^{2P})$ as $\lambda \to 0$. If $\beta_T(\cdot)$ satisfies a Lipschitz condition of order one at $(c_0,\ldots,c_P,d)$ then for all $h \leq H_T = a(\log T)^c$, $a > 0$, $c < \infty$,*

$$N^{1/2}|b_T - \bar{b}_T| = o(N^{-1/2+\epsilon}\log T)\,.$$

Given that $\widehat{b}_{T,B}^{*f}$ in (11) can be made arbitrarily close to $\bar{b}_T$ by taking B sufficiently large, it follows that under the conditions of Theorems 7 and 8 we can expect $\widehat{b}_{T,B}^{*f}$ to closely approximate b_T and, hence, the bootstrap distribution based on the PFSB algorithm (with the sample mean of B bootstrap draws used to represent $E^*[\widehat{d}_T^{*f}]$) to approximate the true finite sampling distribution of $\widehat{d}_T$ to this given order of approximation. This provides the justification for using the PFSB(BA) algorithm to estimate the bias of the statistic of interest and, in turn, produce the bias-adjusted statistic.

Each of the estimators LPR, LPR$_{ba}$, SPLW and SPLW$_{ba}$ is a $\sqrt{N}$ –CAN estimator of d with $N = T^q$ for some $q > 0$. Any one of them can therefore serve as a preliminary estimator of d for the PFSB since the bandwidth is such that $N^{-1/2+\epsilon}\log T = o(1)$, $0 < \epsilon < 0.5$. For the log-periodogram regression estimators,

$$N^{1/2}(\widehat{d}_T - d) \xrightarrow{\mathcal{D}} N\left(0, \frac{\pi^2}{24}\upsilon_P^2\right) , \tag{19}$$

where $N = o(T^{4/5})$ and υ_P^2 gives the variance inflation factor of the estimator. The inflation factor results from the modeling of $\log|\kappa(e^{-\imath\lambda})|^2$ by a polynomial of degree $2P$ that underlies the bias correction. For the local polynomial Whittle estimators,

$$N^{1/2}(\widehat{d}_T - d) \xrightarrow{\mathcal{D}} N\left(0, \frac{1}{4}\upsilon_P^2\right) . \tag{20}$$

For both the LPR and the SPLW estimators the variance inflation factors are $\upsilon_0^2 = 1, \upsilon_1^2 = 2.25, \upsilon_2^2 = 3.52, \upsilon_3^2 = 4.79$ and $\upsilon_4^2 = 6.06$, where $\upsilon_0^2 = 1$ yields the baseline variance for the uncorrected estimator, see Andrews and Guggenberger (2003) and Andrews and Sun (2004). In both cases the optimal MSE bandwidth $N \sim KT^{4/5}$, where K depends on the smoothness of $|\kappa(e^{-\imath\lambda})|^2$ for λ in the vicinity of the origin. In practice the optimal MSE bandwidth seems not to be used much, the values $N = T^{2/5}, T^{1/2}, T^{3/5}$ and $T^{7/10}$ being popular choices.

Note that each of the estimators can, of course, serve as its own preliminary (pre-filtering) value, i.e. setting $d^f = \widehat{d}_T$, or as the preliminary value for another estimator of interest. Furthermore, in the context of the bootstrap algorithm, any bootstrap bias-adjusted version of an initial estimator can serve as a valid pre-filtering value in a subsequent application of the algorithm. This observation, in turn, prompts the following adaptation of the PFSB(BA) algorithm, in which successive bias-adjusted estimators play the role of the preliminary pre-filtering value within an iterative scheme.

4 A Recursive Bias-Correction Procedure

Although the bias of the bias-adjusted estimator $\widetilde{d}_T$ in (12) will be smaller than that of $\widehat{d}_T$, any bias remaining in $E[\widetilde{d}_T] - d$ may still be large because the bias in any preliminary value d^f can be severe in finite samples, and $\widehat{b}^{*f}_{T,B}$ will, likewise, be a biased estimate of its true counterpart b_T. To obtain a more accurate estimate of d we propose a further refinement to the proposed correction of $\widehat{d}_T$ through a recursive algorithm:

BA1′. Initialization: Set $k = 0$ and assign desirable tolerance levels $\tau_1 = \tau_1^{(0)}$ and $\tau_2 = \tau_2^{(0)}$.

For the chosen estimator $\widehat{d}_T$, set $\widetilde{d}_T^{(0)} = \widehat{d}_T$ (**i.e. set** $d^f = \widehat{d}_T$). Now go to Step BA2′.

BA2′. Recursive Calculation: For the kth iteration set the preliminary value of d, namely d^f, to $\widetilde{d}_T^{(k)}$ and repeat Steps BA2 and BA3 of the PFSB(BA) algorithm with $\widehat{d}_T$ therein replaced by $\widetilde{d}_T^{(k)}$ to give, in an obvious notation, $\widetilde{d}_T^{(k+1)} = \widetilde{d}_T^{(k)} - \widetilde{b}_{T,B}^{*f(k)}$. Proceed to Step BA3′.

BA3′. Stopping Rule: If $|\widetilde{d}_T^{(k+1)} - \widetilde{d}_T^{(k)}| > \tau_1$ and $|\widetilde{d}_T^{(0)} - \widetilde{d}_T^{(k)} - \widetilde{b}_{T,B}^{*f(k)}| > \tau_2$ set $k = k+1$, update the tolerance levels $\tau_1 = \tau_1^{(k)}$ and $\tau_2 = \tau_2^{(k)}$, and repeat Step BA2′. Otherwise set $\widetilde{d}_T = \widetilde{d}_T^{(k)}$ and stop.

The rationale behind the recursions is as follows: since the estimator $d^f = \widehat{d}_T$ tends to be severely biased, $\widehat{b}_{T,B}^{*f}$ will on average be a biased estimate of b_T, and the bias adjusted estimate $\widetilde{d}_T$ will therefore still contain some bias. Replacing the initial values $\widehat{d}_T = \widetilde{d}_T^{(0)}$ and $\widehat{b}_{T,B}^{*f} = \widetilde{b}_{T,B}^{*f(0)}$ by $\widetilde{d}_T^{(1)}$ and $\widetilde{b}_{T,B}^{*f(1)}$, and (for general k) $\widetilde{d}_T^{(k-1)}$ and $\widetilde{b}_{T,B}^{*f(k-1)}$ by $\widetilde{d}_T^{(k)}$ and $\widetilde{b}_{T,B}^{*f(k)}$, and so on, produces more accurate estimates and bias assessments. Being based upon more accurate estimators, the updated estimate $\widetilde{d}_T^{(k)}$ would be expected to be closer to the true value of d. The procedure is iterated until no meaningful gain in accuracy is achieved.

To determine if any meaningful gain in accuracy will be achieved by adding a further iteration, two criteria are used. The first, $|\widetilde{d}_T^{(k+1)} - \widetilde{d}_T^{(k)}| > \tau_1^{(k)}$, is based on Cauchy's convergence criterion. Given the stochastic nature of the bias correction mechanism we can think of this as a statistical decision rule in which $\tau_1^{(k)}$ governs the probability of moving from the kth to the $(k+1)$th iteration. Now,

$$|\widetilde{d}_T^{(k+1)} - \widetilde{d}_T^{(k)}| = \left| \widetilde{d}_T^{(k)} - \left(\frac{1}{B} \sum_{b=1}^{B} \widetilde{d}_{T,b}^{*f(k)} \right) \right| ,$$

and since $\widehat{d}_T$ is a $\sqrt{N}$–CAN estimator, given the data and the current and previous bootstrap iterations, $N^{1/2}(\widetilde{d}_{T,b}^{*f(k)} - \widetilde{d}_T^{(k)}) \overset{\mathcal{D}}{\to} N(0, \upsilon^2)$, where $\widetilde{d}_{T,b}^{*f(k)}$ denotes the estimator produced from a bootstrap draw based on the PFSB(BA) algorithm, with $\widetilde{d}_T^{(k)}$ used as the pre-filtering value. This implies that the unconditional asymptotic variance of the difference between successive bias-adjusted estimators in the iterative scheme is given by

$$Var[\widetilde{d}_T^{(k+1)} - \widetilde{d}_T^{(k)}] = Var[\widetilde{d}_T^{(k)}] + \frac{\upsilon^2}{NB} .$$

Furthermore, from the recurrence formula,

$$\widetilde{d}_T^{(k)} = 2\widetilde{d}_T^{(k-1)} - \left(\frac{1}{B} \sum_{b=1}^{B} \widetilde{d}_{T,b}^{*f(k-1)} \right) ,$$

and it follows by the same logic that

$$Var[\widetilde{d}_T^{(k)}] = 4 \cdot Var[\widetilde{d}_T^{(k-1)}] + \frac{\upsilon^2}{NB},$$

where $Var[\widetilde{d}_T^{(1)}] = 4 \cdot Var[\widetilde{d}_T^{(0)}] + \upsilon^2/NB = (4B+1)\upsilon^2/NB$. Moreover, at each iteration the bias adjusted estimate is constructed as a linear combination of asymptotically normal random variables and is itself therefore asymptotically normal. This indicates that $\tau_1^{(k)}$ can be evaluated from percentile points of the normal approximation.

Similarly, the second convergence criterion, $|\widetilde{d}_T^{(0)} - \widetilde{d}_T^{(k)} - \widetilde{b}_{T,B}^{*f(k)}| > \tau_2^{(k)}$, is perhaps best thought of as the decision rule that examines the difference between the current accumulated bias correction, $\widetilde{d}_T^{(0)} - \widetilde{d}_T^{(k)}$, and the current bootstrap estimate of the bias, $\widetilde{b}_{T,B}^{*f(k)}$. From the expression

$$\widetilde{d}_T^{(0)} - \widetilde{d}_T^{(k)} - \widetilde{b}_{T,B}^{*f(k)} = \widetilde{d}_T^{(0)} - \left(\frac{1}{B}\sum_{b=1}^{B} \widetilde{d}_{T,b}^{*f(k)}\right),$$

it follows that the unconditional asymptotic variance,

$$Var[\widetilde{d}_T^{(0)} - \widetilde{d}_T^{(k)} - \widetilde{b}_{T,B}^{*f(k)}] = \frac{(B+1)\upsilon^2}{NB},$$

and the tolerance level $\tau_2^{(k)}$ can once again be set using percentile points from the asymptotic normal approximation.

The interpretation of the convergence criteria as statistical decision rules in which the tolerance levels govern the probability of going from the current to the next iteration suggests that $\tau_1^{(k)}$ and $\tau_2^{(k)}$ be set by reference to conventional critical values used in statistical hypothesis tests. When k is very small we might conjecture that $\widetilde{d}_T^{(k)}$ still contains some bias and we may wish to iterate further unless there is strong evidence that so doing will produce very little change. On the other hand, when k is large the initial estimate $\widetilde{d}_T^{(0)}$ has already undergone several adjustments to produce $\widetilde{d}_T^{(k)}$ and we may prefer to terminate iteration unless there is strong evidence that further iteration will produce additional, substantial correction. We can therefore set $\tau_1^{(k)} = \tau_2^{(k)} = z_{(1-p_k/2)}$ where $G(z_{(1-p)}) = 1 - p$ and p_k, the probability of going from the kth to the $(k+1)$th iteration, is assigned to be large when k is small and vice versa. In the experiments that follow we set $p_0 = 0.95$, $p_1 = 0.9$, and $p_k = (0.1)2^{(1-k)}$ for $k = 2, 3, \ldots$ for uncorrected LPR and SPLW; and $p_0 = 0.9$, $p_k = (0.1)2^{-k}$ for $k = 1, 2, 3, \ldots$ for LPR-BA and SPLW-BA with $P = 1$. We comment further on the appropriate choice of the p_k when discussing our experimental results below.

5 Simulation Exercise

5.1 Simulation Design

In this section we illustrate the performance of the bootstrap bias-corrected estimators via a small simulation experiment. Following Andrews and Guggenberger (2003) we simulate data from a Gaussian $ARFIMA(1, d, 0)$ process,

$$(1 - L)^d \Phi(z) y(t) = \varepsilon(t)\,,\; 0 < d < 0.5\,, \tag{21}$$

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 choice of this model is motivated, in part, by earlier work that highlights the distinct finite sample bias of the LPR estimator of d in this setting, when the value of ϕ is positive and large (See Agiakloglou et al., 1993). Indeed, Andrews and Guggenberger (2003) document substantial remaining bias in the bias-corrected version of the LPR estimator in the presence of a large autoregressive parameter. That is, the impetus for applying bootstrap-based bias corrections to the various estimators is particularly strong in this setting.

The process in (21) 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 $ARFIMA(p, d, q)$ process and the generated pseudo-random $\varepsilon(t)$ (see, for instance, Brockwell and Davis, 1991, §5.2). The ARFIMA ACF for given T, ϕ, θ, and d is calculated using Sowell's (1992) algorithm as modified by Doornik and Ooms (2001).

The estimators that we bias correct via the iterative PFSB(BA) algorithm are: LPR, LPR$_{ba}$ ($P = 1, 2$), SPLW and SPLW$_{ba}$ ($P = 1, 2$), as described above and implemented with fixed bandwidth[4] $N = T^{0.7}$ and $B = 1000$. The order (h) of the autoregressive approximation underlying the sieve component of the bootstrap algorithm is chosen via AIC, and Burg's algorithm is used to estimate the autoregressive parameters.

Based on the R replications, for each estimator of d, we report the bias and mean square error (MSE). For comparative purposes, we also document the performance of the unadjusted (LPR, SPLW) estimators and the estimators that are analytically adjusted (LPR$_{ba}$ and SPLW$_{ba}$; $P = 1, 2$). That is, we are interested, in particular, in documenting: 1) any improvement that can be had by using the bootstrap method *rather* than an analytical

[4] $N = T^{0.7}$ was chosen so as to satisfy the minimum requirement for asymptotic normality of LPR (Robinson, 1995b)

method to bias correct a given estimator; and 2) any *additional* improvement associated with bias-correcting (via the bootstrap) an estimator that has already been bias corrected via analytical means.

For each estimator considered (i.e. each of the two base estimators, LPR and SPLW, and all of the analytically and bootrap bias-corrected versions thereof), we also document the empirical coverage (over the Monte Carlo replications) of the nominal 95% confidence intervals, plus the average length of the given intervals. The 95% confidence intervals (CIs) are constructed from each of the R bootstrap distributions (each, in turn, based on B bootstrap draws) as: $\left\{\widetilde{d}_T(L), \widetilde{d}_T(U)\right\}$, where $\widetilde{d}_T(L)$ ($\widetilde{d}_T(U)$) denotes the lower (upper) bound of a highest density interval, in which the narrowest interval with 95% coverage for the bootstrap distribution is selected. For any given estimator $\widetilde{d}_T$, empirical coverage for each interval type is calculated as the proportion of times (in R replications) that each interval covers the true value of d. The average length of each interval (across the R replications) is also recorded. These coverage and length statistics for the bootstrap-based estimators are compared with the empirical coverage and (constant) length of 95% intervals constructed for the unadjusted (or analytically-adjusted) LPR and SPLW estimators, as based on the appropriate asymptotic distributions, in (19) and (20) respectively. Note that the value of B used here implies, from the Dvoretsky–Kiefer–Wolfowitz inequality, that $P(\sup_x |\overline{F}^*_{\widetilde{d}_T,B}(x) - F^*_{\widetilde{d}_T}(x)| > \delta) < 2\exp(-\delta^2(1000))$, where $\overline{F}^*_{\widetilde{d}_T,B}(x)$ is the empirical (bootstrap) distribution of $\widetilde{d}_T$, based on B bootstrap draws, and $F^*_{\widetilde{d}_T}(x)$ is the distribution of $\widetilde{d}_T$ under the probability law induced by the bootstrap. The bootstrap distributions are evaluated using a kernel density estimate based on a Gaussian kernel with bandwidth equal to 75% of the over-smoothed bandwidth, that is, $0.75\, s \sqrt[5]{(243/35B)}$ where s is the standard deviation calculated from the B data values being smoothed, see Wand and Jones (1995).

We record results for the bootstrap-based estimators produced through formal application of the stopping rules described above. To the two stochastic stopping criteria we add a deterministic criterion, whereby the iterative scheme ceases if $\widetilde{d}_T^{(k+1)} < -1$ or ≥ 1.5 and the estimator $\widetilde{d}_T^{(k)}$ retained as the final choice. We also record results for the estimators based on only 1 and 2 iterations of the iterative method ($k = 1, 2$ in Steps BA2′ and BA3′). The following section records all numerical results associated with the LPR estimator, and Section 5.3 all results for the SPLW estimator. Note that most results for $T = 200$ and $d = 0.3$ are omitted for brevity. The coverage and length results for the three different values of ϕ are reported after averaging over all four values of d, including $d = 0.3$.

5.2 Simulation Results: LPR

Tables 1 and 2 record (for $T = 100$ and 500 respectively) the bias and MSE results for all estimators based on the LPR method, with all acronyms as defined in the Introduction. The columns headed 'SSR' report the results based on the stochastic stopping rules discussed in Section 4 and detailed at the end of Section 5.1. Table 3 summarizes the empirical coverage performance of highest probability density (HPD) confidence intervals for the alternative estimators, for both sample sizes and based on a nominal coverage of 95%. The second panel of this table records the average length (across simulations) of the 95% HPD intervals for all cases. Coverage (and length) results for the nominal level of 90% are qualitatively similar, and hence are not reported. In all tables the most favorable result for each parameter setting is highlighted in bold.

The key message to be gleaned from the numerical results is that the bootstrap technique *does* reduce bias, but with the most substantial gains to be had by using the bootstrap algorithm to bias-adjust an estimator that has already been bias adjusted analytically. For example, for $T = 100$, and for $\phi = 0.3, 0.6$, in all but one of the 6 cases, the smallest bias is produced by bias adjusting (via the bootstrap) the LPR-BA ($P = 2$) estimator once, with no subsequent iteration ($k = 0$). For $T = 500$, the LPR-BA ($P = 2$) estimator, bootstrap-bias-adjusted once, is the least biased estimator for *all* 3 values of d and for $\phi = 0.3, 0.6$. Importantly, for these two values of ϕ (and for both sample sizes) if one compares the MSE of the LPR-BA ($P = 2$) estimator adjusted once by the bootstrap, with that of LPR-BA ($P = 3$), the reduction in bias produced by the bootstrap technique is *not* obtained at the expense of MSE, with the two estimators having very similar MSE's, and one not systematically dominating the other in terms of this performance measure. For $\phi = 0.9$, *all* versions of the LPR estimator, including the bootstrap bias adjusted versions, are very biased. That said, for $T = 500$, the estimator with the *smallest* bias is the raw LPR ($P = 0$) estimator bootstrap bias-adjusted three times ($k = 2$).

In terms of coverage, a combination of analytical and bootstrap-based bias adjustments once again yields the best results overall, with either LPR-BA$_{sb}$ ($P = 1, k = 1$) (i.e., LPR-BA ($P = 1$) bootstrapped bias-adjusted twice) or LPR-BA$_{sb}$ ($P = 2, k = 0$) (i.e., LPR-BA ($P = 2$) bootstrapped bias-adjusted once) having the best empirical coverage – and very *accurate* empirical coverage – in all four cases recorded in Table 3 for $\phi = 0.3, 0.6$. Once again, all coverage results for $\phi = 0.9$ are poor, although, for what it is worth, for $T = 500$, the bootstrapped bias-adjusted LPR-BA ($P = 2$) produces the most accurate coverage interval (at 32%!).

In terms of the length of the 95% intervals, there are two key points to note. Firstly, it is the asymptotic intervals which are the most narrow, but this precision is at the expense of very inaccurate coverage. Secondly, the coverage accuracy yielded by the bootstrap is *not* at the expense of precision. That is, any bootstrap-based bias correction that improves coverage produces a negligible change in the length of the interval. This result provides an interesting contrast with the corresponding results for analytical bias-adjustment; i.e. any such analytical adjustment that improves coverage does so at the expense of a decrease in precision, with the 95% intervals widening as the value of P increases.

This raises the question of how the sieve bootstrap is able to bias correct the LPR or a LPR-BA estimator without incurring any loss of precision. The motivation underlying log periodogram regression is that

$$\frac{I_T(\lambda)2\pi|1-e^{-\imath\lambda}|^{2d}}{\sigma^2|\kappa(e^{\imath\lambda})|^2} \xrightarrow{\mathcal{D}} Exp(1) \tag{22}$$

and using the approximation $|1-e^{-\imath\lambda}|^{2d} = |\lambda|^{2d}(1+o(1))$ as $\lambda \to 0$ we have

$$\log(I_T(\lambda_j)) = \alpha - 2d\log(\lambda_j) + \eta_j \tag{23}$$

where $E[\eta_j] = 0$ and

$$\begin{aligned} \alpha &= \log|\kappa(1)|^2 + \log\left(\frac{|\kappa(e^{\imath\lambda_j})|^2}{|\kappa(1)|^2}\right) + \log\left(\frac{\lambda_j^{2d}}{|1-e^{-\imath\lambda_j}|^{2d}}\right) + C \\ &\approx \text{const. as} \quad \lambda_j = 2\pi j/T \to 0\,. \end{aligned} \tag{24}$$

The analytical correction replaces the simple regression in (23), where α is treated as a constant, by the multiple regression

$$\log(I_T(\lambda_j)) = \sum_{p=0}^{P} \alpha_p \lambda_j^{2p} - 2d\log(\lambda_j) + \eta_j\,, \tag{25}$$

the rationale being that the term $\sum_{p=0}^{P} \alpha_p \lambda_j^{2p}$ provides a better approximation to the Maclaurin series expansion of (24) than supposing it is constant in a neighbourhood of zero. The introduction of λ_j^{2p}, $p = 1,\ldots,P$, in (25) reduces the bias in the estimate of d, but it is also the presence of these additional regressors that causes the variance inflation seen in (19).

The PFSB, on the other hand, takes the specification of the regression in (23) or (25) as given and adjusts the estimator by mimicking the sampling behaviour of the regressand. Recall that $I_T(\lambda) = (2\pi)^{-1}\sum_{r=1-T}^{T-1}\widehat{\gamma}(r)e^{\imath\lambda r}$. Hosking (1996) shows that when d is large the $\widehat{\gamma}(r)$

have substantial negative bias relative to the true autocovariances, even for moderate to large samples. The PFSB reduces the memory in the "data" to which the sieve bootstrap is applied, via the pre–filtering procedure, so as to give a near optimal convergence rate when implicitly assessing the corresponding bias in $\log(I_T(\lambda))$. Whether it is applied to (23) or (25), the PFSB is thereby able to attack the problem of bias in the estimation of d without compromising the pivotal nature of the ratio in (22), the basic result that underlies the log periodogram regressions and determines the estimators' variance.

5.3 Simulation Results: SPLW

Tables 4 and 5 record (for $T = 100$ and 500 respectively) the bias and MSE results for all estimators based on the SPLW method (with all acronyms as defined in the Introduction), whilst Table 6 records the 95% interval coverage and length statistics, for all cases. Once again, the most favorable result for each parameter setting is highlighted in bold in all tables. As with the LPR-based estimators, the bootstrap-based bias adjustment yields the largest bias reductions, but only when applied to an SPLW estimator that has already been analytically bias adjusted. In contrast with the LPR-based results, these bias gains are evident only for the larger of the two sample sizes ($T = 500$), with there being no gain (over full analytical adjustments) in the $T = 100$ case. The bias gains (for the $T = 500$ case) are for $\phi = 0.3, 0.6$ only, with the least biased estimator for $\phi = 0.9$ being the SPLW-BA ($P = 3$) estimator. The biases of all SPLW-based estimators are similar to the biases of the comparable LPR-based estimators, and as with the LPR-based estimators, the reduction in bias produced by the bootstrap technique (in certain cases) is not obtained at the expense of MSE. Interestingly enough, although the use of a stochastic stopping rule is appealing, as was the case for the LPR results, it does not guarantee an improvement in performance over using a fixed number of iterations.

The coverage results for the SPLW-based estimators are qualitatively identical to those for the LPR case; in particular, the bootstrap bias adjustment of an already analytically adjusted estimator yields the best coverage for $\phi = 0.3, 0.6$, for both sample sizes - and very accurate coverage at that. Although the bootstrapped bias-adjustment of LPR-BA ($P = 2$) produces the most accurate coverage for the $\phi = 0.9$ case (for both sample sizes), the coverage results are poor for all estimators in this part of the parameter space. Once again, the bootstrap-based bias adjustment is not accompanied by an increase in interval length, in contrast with the analytical bias adjustment. As a consequence, the bootstrap method can be used to yield coverage that is close to the nominal level without sacrificing inferential precision.

6 Conclusion

This paper has developed a bootstrap method for bias correcting semi–parametric estimators of the long memory parameter in fractionally integrated processes. The method involves applying the sieve bootstrap to data pre-filtered by a preliminary semi-parametric estimate of the long memory parameter. In addition to providing theoretical (asymptotic) justification for using the bootstrap techniques, we document the results of simulation experiments, in which the finite sample performance of the (bias-adjusted) estimators is compared with that of both unadjusted estimators and estimators adjusted via analytical means. The numerical results are very encouraging, and suggest that the bootstrap bias correction *can* yield more accurate inferences about long memory dynamics in the types of samples that are encountered in practice.

Appendix A: Proofs

Proof of Theorem 3: 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). □

Proof of Lemma 4: Let $\phi_h^f(z) = \sum_{j=0}^h \phi_h^f(j)z^j$ where $\phi_h^f(1), \ldots, \phi_h^f(h)$ denote the coefficients in the MMSEP(h) of the process

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

and let σ_h^{f2} denote the MMSE. Set $\kappa^f(z) = \kappa(z)/(1-z)^{d-d^f}$ and define $\kappa_h^f(z) = \{\phi_h^f(z)\}^{-1}$ by replacing the coefficients of $\bar{\phi}_h(z)$ by those of $\phi_h^f(z)$ in the recursions in equation (14). Then, trivially,

$$|\bar{\kappa}_h(e^{\imath\lambda})|^2 - |\kappa(e^{\imath\lambda})|^2 = (|\bar{\kappa}_h(e^{\imath\lambda})|^2 - |\kappa_h^f(e^{\imath\lambda})|^2) + (|\kappa_h^f(e^{\imath\lambda})|^2 - |\kappa^f(e^{\imath\lambda})|^2) + (|\kappa^f(e^{\imath\lambda})|^2 - |\kappa(e^{\imath\lambda})|^2)\,. \tag{26}$$

Consider the first term in (26), $|\bar{\kappa}_h(e^{\imath\lambda})|^2 - |\kappa_h^f(e^{\imath\lambda})|^2$. By definition

$$\bar{\kappa}_h(z) - \kappa_h^f(z) = \frac{\phi_h^f(z) - \bar{\phi}_h(z)}{\bar{\phi}_h(z)\phi_h^f(z)}\,,$$

and since $\bar{\phi}_h(z) \neq 0$ and $\phi_h^f(z) \neq 0$, $|z| \leq 1$, there exists an $\epsilon > 0$ such that

$$\begin{aligned}
|\bar{\kappa}_h(z) - \kappa_h^f(z)| &\leq \epsilon^{-2}|\phi_h^f(z) - \bar{\phi}_h(z)| \\
&\leq \epsilon^{-2}\sum_{j=0}^{h}|\phi_h^f(j) - \bar{\phi}_h(j)| \quad \text{for all} \quad |z| \leq 1\,.
\end{aligned}$$

But

$$\begin{aligned}
\sum_{j=0}^{h}|\phi_h^f(j) - \bar{\phi}_h(j)| &\leq \left(h\sum_{j=0}^{h}|\phi_h^f(j) - \bar{\phi}_h(j)|^2\right)^{1/2} \\
&= O\left(h\left(\frac{\log T}{T}\right)^{1/2(1-2\max\{0,d-d^f\})}\right) \\
&= \exp(\delta_T(\log T - \log\log T))O\left(h\left(\frac{\log T}{T}\right)^{1/2}\right)
\end{aligned}$$

by Theorem 3 and the fact that $|d^f - d| < \delta_T$ by assumption. It follows that $|\bar{\kappa}_h(e^{\imath\lambda}) - \kappa_h^f(e^{\imath\lambda})| \leq \nu_{1,T}$ uniformly in λ, and hence $\left||\bar{\kappa}_h(e^{\imath\lambda})|^2 - |\kappa_h^f(e^{\imath\lambda})|^2\right| \leq \nu_{1,T}$ uniformly in λ. We can therefore interchange limit operations (Apostol, 1960, Theorem 13.3) to give

$$\lim_{T\to\infty}\lim_{\lambda\to 0}\left||\bar{\kappa}_h(e^{\imath\lambda})|^2 - |\kappa_h^f(e^{\imath\lambda})|^2\right| = \lim_{\lambda\to 0}\lim_{T\to\infty}\left||\bar{\kappa}_h(e^{\imath\lambda})|^2 - |\kappa_h^f(e^{\imath\lambda})|^2\right| \leq \nu_{1,T}\,.$$

For the second term in (26), $|\kappa_h^f(\rho()|^2 - |\kappa^f(e^{\imath\lambda})|^2$, we have

$$\kappa_h^f(z) - \kappa^f(z) = \frac{1 - \kappa^f(z)\phi_h^f(z)}{\phi_h^f(z)}\,,$$

giving us the bound

$$|\kappa_h^f(z) - \kappa^f(z)| \leq \epsilon^{-1}|1 - \kappa^f(z)\phi_h^f(z)| \quad \text{for all} \quad |z| \leq 1\,.$$

Let $\rho_h(z) = \sum_{j\geq 1}\rho_h(j)z^j = 1 - \kappa^f(z)\phi_h^f(z)$. Then from Parseval's relation

$$\sum_{j\geq 1}\rho_h(j)^2 = \int_{-\pi}^{\pi}|1 - \kappa^f(e^{\imath\lambda})\phi_h^f(e^{\imath\lambda})|^2 d\lambda = 2\pi\sigma^{-2}(\sigma_h^{f2} - \sigma^2)$$

and from the Levinson–Durbin recursions (Durbin, 1960; Levinson, 1947) we have $\sigma_h^{f2} = (1-\phi_h^f(h)^2)\sigma_{h-1}^{f2}$. Substituting sequentially in the recurrence formula $\sigma_h^{f2} = \sigma_{h+1}^{f2} + \phi_h^f(h)^2\sigma_h^{f2}$

leads to the series expansion $\sigma_h^{f2} - \sigma^2 = \sum_{r=h}^{\infty} \phi_r^f(r)^2 \sigma_r^{f2}$, from which we obtain the bound

$$\sum_{j \geq 1} \rho_h(j)^2 \leq 2\pi\sigma^{-2} E[w^f(t)^2] \sum_{r=h}^{\infty} \phi_r^f(r)^2 .$$

But $\phi_h^f(h) \sim |d - d^f|/h$ as $h \to \infty$ (Inoue, 2002; Inoue and Kasahara, 2004) and therefore we can infer that

$$\sum_{j \geq 1} \rho_h(j)^2 \leq \text{const.} \frac{|d - d^f|^2}{h^{2|d|}} \zeta(2(1 - |d|)),$$

where $\zeta(\cdot)$ denotes the Riemann zeta function. It follows that $\lim_{h\to\infty} \rho_h(e^{\imath\lambda}) = 0$ and that $\lim_{T\to\infty} |\rho_h(e^{\imath\lambda})|^2 = O(\delta_T^2 h^{-2|d|})$ almost everywhere on $[-\pi, \pi]$. Hence we can conclude that $\lim_{T\to\infty} \lim_{\lambda\to 0} \left| |\kappa_h^f(e^{\imath\lambda})|^2 - |\kappa^f(e^{\imath\lambda})|^2 \right| \leq \nu_{2,T}$.

The third and final term in (26) is

$$|\kappa^f(e^{\imath\lambda})|^2 - |\kappa(e^{\imath\lambda})|^2 = |\kappa(e^{\imath\lambda})|^2 (|1 - e^{\imath\lambda}|^{2(d^f - d)} - 1) . \tag{27}$$

Substituting $|1 - e^{\imath\lambda}|^{2(d^f - d)} = \exp(2(d^f - d) \log |1 - e^{\imath\lambda}|)$ into (27) and using the expansions $|\exp(x) - 1| = |x|(1 + o(1))$ for x in a neighbourhood of the origin and $\log |1 - e^{\imath\lambda}| = \log |2 \sin(\lambda/2)| = \log |\lambda| + \log(1 + o(|\lambda|))$ as $\lambda \to 0$ we can conclude, via the squeezing principle, that

$$\left| |\kappa(e^{\imath\lambda})|^2 (|1 - e^{\imath\lambda}|^{2(d^f - d)} - 1) \right| \leq 2 \sup_{[-\pi,\pi]} |\kappa(e^{\imath\lambda})|^2 |d^f - d| (\log 2\pi N + \log T)$$

for all $\lambda \in [2\pi/T, 2\pi N/T]$ as $T \to \infty$. We can therefore infer that (27) is $O(\delta_T \log T)$ or smaller, uniformly in λ for all $\lambda \in [2\pi/T, 2\pi N/T]$. The lemma now follows. □

Proof of Theorem 7: Substituting

$$P^*\{N^{1/2}(\widehat{d}_T^{*f} - d_T^f) < x + \bar{b}_T\} = G((x + N^{1/2}\bar{b}_T)/\upsilon) + o(N^{-1/2})$$

and

$$P\{N^{1/2}(\widehat{d}_T - d) < x + b_T\} = G((x + N^{1/2}b_T)/\upsilon) + o(N^{-1/2})$$

into equations (17) and (18), subtracting, and using the triangular inequality, it follows that $|P^*\{N^{1/2}(\widehat{d}_T^{*f} - E^*[\widehat{d}_T^{*f}]) < x\} - P\{N^{1/2}(\widehat{d}_T - E[\widehat{d}_T]) < x\}|$ is less than or equal to

$$|G((x + N^{1/2}b_T)/\upsilon) - G((x + N^{1/2}\bar{b}_T)/\upsilon)| + o(N^{-1/2}) .$$

But

$$\sup_x |G((x+N^{1/2}b_T)/\upsilon) - G((x+N^{1/2}\bar{b}_T)/\upsilon)| \leq \frac{N^{1/2}}{\upsilon\sqrt{2\pi}}|b_T - \bar{b}_T|$$

by the first mean value theorem for integrals(Apostol, 1960, Theorem 7.30). □

Proof of Theorem 8: If $\beta_T(\cdot)$ is Lipschitz of order one at $(c_0, \ldots, c_P, d)$ then there exists a constant $K < \infty$ such that for $(\bar{c}_0, \ldots, \bar{c}_P, d_T^f)$ in a neighbourhood of $(c_0, \ldots, c_P, d)$

$$|\beta_T(c_0, \ldots, c_P, d) - \beta_T(\bar{c}_0, \ldots, \bar{c}_P, d_T^f)| \leq K \left(\sum_{p=0}^{P} |c_p - \bar{c}_p|^2 + |d_T^f - d|^2 \right)^{1/2} . \quad (28)$$

An immediate consequence of Lemma 4, however, is that $|c_p - \bar{c}_p| = \nu_{1,T} + \nu_{2,T} + \nu_{3,T}$ for $p = 0, \ldots, P$. Moreover, $|d_T^f - d| = o(N^{-1/2+\epsilon})$ – recall the discussion preceeding Theorem 6 – implying that we can set $\delta_T = o(N^{-1/2+\epsilon})$, and $N^{-1/2+\epsilon}\log T = o(1)$, in which case $\nu_{1,T} = \exp(\delta_T(\log T - \log\log T))O(h(\log T/T)^{1/2}) = o(N^{-1/2+\epsilon})$, $\nu_{2,T} = o(N^{-1/2+\epsilon}h^{-|d|})$ and $\nu_{3,T} = o(N^{-1/2+\epsilon}\log T)$. Extracting the dominant term on the right hand side in (28) gives the desired result. □

Appendix B: Tables

Table 1: Bias and mean square error (MSE) for all LPR-based estimators: $T = 100$. Unadjusted (LPR); analytically bias-adjusted (LPR-BA); bootstrap bias-adjusted (LPR_{sb} for $k = 0, 1, 2$); bootstrap bias-adjusted *after* analytical adjustment (LPR-BA_{sb}). The lowest bias (in absolute value) and MSE for each parameter setting are highlighted in bold.

			LPR-BA			LPR_{sb}				$LPR\text{-}BA_{sb}$			
										$P=1$			$P=2$
		LPR	$P=1$	$P=2$	$P=3$	$k=0$	$k=1$	$k=2$	SSR	$k=0$	$k=1$	SSR	$k=0$
d	ϕ	Bias											
0	0.3	0.1445	0.0366	0.0138	0.0236	0.1255	0.0944	0.0368	0.0798	0.0165	**-0.0063**	0.0108	-0.0161
	0.6	0.3947	0.2	0.1039	0.0725	0.3511	0.2799	0.1506	0.2655	0.1574	0.0919	0.1485	**0.0564**
	0.9	0.823	0.7402	0.654	**0.5969**	0.8	0.7188	0.8053	0.7439	0.7031	0.6234	0.6915	0.6161
0.2	0.3	0.14	0.0395	0.0161	0.0262	0.1207	0.0886	0.0252	0.0769	0.022	-0.0116	0.0152	**-0.0093**
	0.6	0.3887	0.2017	0.1047	0.0746	0.3401	0.2609	0.1183	0.2459	0.1549	0.0805	0.1474	**0.0601**
	0.9	0.7968	0.731	0.6558	**0.5937**	0.818	0.8612	0.9425	0.89	0.7301	0.6253	0.7162	0.6534
0.4	0.3	0.1374	0.0461	0.0194	0.0309	0.111	0.0684	-0.013	0.059	0.0229	-0.0193	0.0127	**-0.0047**
	0.6	0.378	0.2051	0.1063	0.073	0.3319	0.2555	0.1178	0.2349	0.1546	0.0713	0.1368	**0.062**
	0.9	0.7245	0.691	0.6333	**0.5706**	0.8107	0.9839	1.2222	1.1407	0.7485	0.8018	0.7676	0.6859
d	ϕ	MSE											
0	0.3	**0.0463**	0.0753	0.1483	0.2369	0.065	0.1396	0.3867	0.1869	0.1349	0.2549	0.1602	0.2525
	0.6	0.181	**0.1125**	0.1543	0.2348	0.1711	0.2041	0.4095	0.2461	0.1515	0.2807	0.1841	0.2483
	0.9	0.7031	0.6189	**0.5658**	0.5861	0.6948	0.7188	0.8053	0.7552	0.6235	0.6697	0.6471	0.6341
0.2	0.3	**0.0449**	0.0737	0.1409	0.2247	0.0612	0.1276	0.3602	0.1664	0.1187	0.2478	0.1532	0.222
	0.6	0.1765	**0.1117**	0.1493	0.231	0.164	0.1942	0.4002	0.2357	0.1422	0.2664	0.162	0.2301
	0.9	0.6589	0.6026	**0.562**	0.5752	0.7375	0.9677	1.6844	1.2435	0.6903	0.7542	0.7804	0.718
0.4	0.3	**0.044**	0.0747	0.1415	0.2372	0.0616	0.1201	0.3477	0.1507	0.1174	0.2565	0.1585	0.2253
	0.6	0.1676	**0.1135**	0.1498	0.2403	0.1635	0.2106	0.4723	0.2811	0.1486	0.2976	0.1944	0.242
	0.9	0.5519	0.5458	**0.5325**	0.5532	0.7585	1.3521	2.8938	2.3084	0.7647	1.2723	1.0173	0.816

Table 2: Bias and mean square error (MSE) for all LPR-based estimators: $T = 500$. Unadjusted (LPR); analytically bias-adjusted (LPR-BA); bootstrap bias-adjusted (LPR_{sb} for $k = 0, 1, 2$); bootstrap bias-adjusted *after* analytical adjustment (LPR-BA_{sb}). The lowest bias (in absolute value) and MSE for each parameter setting are highlighted in bold.

			$LPR\text{-}BA$			LPR_{sb}				$LPR\text{-}BA_{sb}$			
										$P=1$			$P=2$
		LPR	$P=1$	$P=2$	$P=3$	$k=0$	$k=1$	$k=2$	SSR	$k=0$	$k=1$	SSR	$k=0$
d	ϕ	Bias											
0	0.3	0.0619	0.0097	0.006	0.0026	0.0351	-0.002	-0.0607	-0.0053	0.0025	-0.0089	0.0018	**0.0001**
	0.6	0.2221	0.0671	0.0244	0.009	0.1603	0.0652	-0.1	0.0446	0.0282	-0.0323	0.0168	**-0.0016**
	0.9	0.6736	0.4946	0.3707	0.2814	0.5927	0.4628	**0.2351**	0.4014	0.4114	0.2818	0.3802	0.2917
0.2	0.3	0.0601	0.0101	0.0066	0.0036	0.033	-0.0044	-0.0642	-0.0063	0.002	-0.0108	0.0019	**-0.0014**
	0.6	0.2205	0.0679	0.0253	0.0105	0.1561	0.057	-0.114	0.0344	0.027	-0.0353	0.0166	**-0.0027**
	0.9	0.6691	0.4948	0.3713	0.284	0.5972	0.4758	**0.261**	0.4168	0.4045	0.266	0.3765	0.2842
0.4	0.3	0.0613	0.0151	0.0126	0.011	0.032	-0.0079	-0.072	-0.0087	0.0034	-0.0126	0.0049	**0**
	0.6	0.2206	0.0725	0.0304	0.0174	0.1488	0.0392	-0.1489	0.0116	0.0262	-0.0418	0.019	**-0.0041**
	0.9	0.6534	0.4908	0.3704	**0.2856**	0.6621	0.6785	0.6175	0.6529	0.4227	0.3126	0.3958	0.2876
d	ϕ	MSE											
0	0.3	**0.0103**	0.0165	0.0293	0.0409	0.0131	0.0271	0.0783	0.036	0.0236	0.0414	0.0284	0.0389
	0.6	0.0558	**0.021**	0.0302	0.0413	0.0385	0.04	0.1237	0.0817	0.0288	0.0624	0.0554	0.0463
	0.9	0.4603	0.2614	0.1675	**0.1208**	0.3636	0.2468	0.1611	0.3356	0.1999	0.1549	0.2361	0.1393
0.2	0.3	**0.0102**	0.0168	0.0303	0.042	0.013	0.0272	0.0782	0.031	0.0235	0.0404	0.0307	0.0387
	0.6	0.0552	**0.0213**	0.0307	0.0416	0.0371	0.0382	0.1253	0.0798	0.0288	0.062	0.0511	0.0456
	0.9	0.4542	0.2614	0.1675	**0.1221**	0.3715	0.2672	0.198	0.3487	0.1941	0.1432	0.2369	0.1334
0.4	0.3	**0.0103**	0.0169	0.0303	0.0415	0.0127	0.0261	0.0748	0.0274	0.0219	0.0351	0.0265	0.0356
	0.6	0.0552	**0.0219**	0.0312	0.042	0.0345	0.0349	0.1288	0.0953	0.0278	0.0587	0.0446	0.043
	0.9	0.4342	0.2579	0.1678	**0.1243**	0.4631	0.5524	0.7172	0.5893	0.218	0.2031	0.2479	0.1446

Table 3: Empirical coverage and length of (nominal 95%) HPD intervals for all LPR-based estimators: $T = 100, 500$. Unadjusted (LPR); analytically bias-adjusted (LPR-BA); bootstrap bias-adjusted (LPR$_{sb}$ for $k = 0, 1, 2$); bootstrap bias-adjusted *after* analytical adjustment (LPR-BA$_{sb}$). Figures are averaged over all values of d used in the experimental design for each value of ϕ. Coverages for the intervals based on the asymptotic distribution of the LPR and analytically bias-adjusted (LPR-BA) estimators are also reported for comparison. The empirical coverage closest to the nominal 95%, and the shortest length, are highlighted in bold.

								$LPR\text{-}BA_{sb}$			*Asymptotic interval*		
			$LPR\text{-}BA$			LPR_{sb}		$P=1$		$P=2$		$LPR\text{-}BA$	
		LPR	$P=1$	$P=2$	$k=0$	$k=1$	$k=2$	$k=0$	$k=1$	$k=0$	LPR	$P=1$	$P=2$
ϕ	T	Coverage											
0.3	100	0.9015	0.9795	0.973	0.9048	0.888	0.841	0.9773	**0.9612**	0.9635	0.7563	0.8408	0.8035
	500	0.8793	0.9748	0.9698	0.912	0.9128	0.9075	0.9683	**0.9555**	0.9703	0.8343	0.9083	0.8873
0.6	100	0.2058	0.916	0.9713	0.2475	0.301	0.3328	0.9248	0.9092	**0.9595**	0.1918	0.7078	0.786
	500	0.0698	0.9388	0.971	0.0898	0.159	0.2155	0.9435	**0.944**	0.9738	0.1593	0.8565	0.884
0.9	100	0	0.1568	**0.5945**	0.0005	0.0013	0.0195	0.1898	0.2405	0.588	0.001	0.102	0.3065
	500	0	0.003	0.215	0	0	0.0005	0.0063	0.014	**0.3168**	0	0.02	0.267
ϕ	T	Interval length											
0.3	100	**0.6413**	1.1082	1.5664	0.6425	0.6452	0.6507	1.107	1.0978	1.5542	0.5016	0.7523	0.9404
	500	0.3278	0.5267	0.6976	**0.3275**	0.3279	0.3303	0.5271	0.5275	0.6984	0.2856	0.4283	0.5354
0.6	100	0.6404	1.1046	1.5622	0.6409	0.641	**0.6392**	1.1045	1.0931	1.5492	0.5016	0.7523	0.9404
	500	0.3308	0.5274	0.6983	0.3294	**0.329**	0.3306	0.5269	0.5273	0.6989	0.2856	0.4283	0.5354
0.9	100	0.6114	1.0347	1.4638	0.6056	0.5716	**0.5062**	0.9663	0.9251	1.3499	0.5016	0.7523	0.9404
	500	0.3306	0.5224	0.6954	0.3325	0.3252	**0.315**	0.5241	0.5244	0.6957	0.2856	0.4283	0.5354

Table 4: Bias and mean square error (MSE) for all LPW-based estimators: $T = 100$. Unadjusted (LPW); analytically bias-adjusted (LPW-BA); bootstrap bias-adjusted (LPW_{sb} for $k = 0, 1, 2$); bootstrap bias-adjusted *after* analytical adjustment ($LPW\text{-}BA_{sb}$). The lowest bias (in absolute value) and MSE for each parameter setting are highlighted in bold.

			$SPLW\text{-}BA$			$SPLW_{sb}$				$SPLW\text{-}BA_{sb}$			
										$P=1$			$P=2$
		$SPLW$	$P=1$	$P=2$	$P=3$	$k=0$	$k=1$	$k=2$	SSR	$k=0$	$k=1$	SSR	$k=0$
d	ϕ	Bias											
0	0.3	0.1327	**-0.0064**	-0.0393	-0.0715	0.1191	0.0997	0.0647	0.1003	0.0111	0.0315	0.0078	-0.025
	0.6	0.3993	0.1629	0.053	**-0.0214**	0.3697	0.3243	0.2456	0.3252	0.153	0.1328	0.1492	0.0504
	0.9	0.8239	0.7139	0.6192	**0.5165**	0.8164	0.8035	0.7706	0.8043	0.7125	0.7044	0.7097	0.6209
0.2	0.3	0.1268	**-0.0058**	-0.0397	-0.0709	0.1127	0.0928	0.0572	0.0929	0.0119	0.0311	0.0084	-0.0216
	0.6	0.3922	0.1633	0.0538	**-0.0214**	0.3586	0.3062	0.2133	0.3072	0.1494	0.1228	0.1469	0.0535
	0.9	0.7997	0.7029	0.6154	**0.5068**	0.8296	0.8687	0.8438	0.8755	0.7227	0.7113	0.7207	0.6378
0.4	0.3	0.1246	**0.0004**	-0.034	-0.0668	0.1081	0.0842	0.0395	0.0846	0.0129	0.0234	0.0109	-0.0141
	0.6	0.3831	0.1668	0.0586	**-0.0193**	0.3534	0.3035	0.206	0.3039	0.1466	0.1124	0.146	0.0565
	0.9	0.7363	0.6724	0.5942	**0.4913**	0.8291	0.8785	0.7524	0.9266	0.7419	0.6804	0.7288	0.6583
d	ϕ	MSE											
0	0.3	**0.0352**	0.0523	0.1128	0.1993	0.0393	0.0624	0.1518	0.0623	0.083	0.1541	0.0896	0.179
	0.6	0.1787	**0.0789**	0.1129	0.1921	0.1621	0.1556	0.1999	0.1562	0.1008	0.172	0.107	0.1777
	0.9	0.6969	0.562	0.4913	**0.4533**	0.6973	0.7108	0.7566	0.7116	0.5869	0.6475	0.5835	0.5553
0.2	0.3	**0.0339**	0.0522	0.1104	0.1944	0.0379	0.0605	0.1479	0.0602	0.0766	0.1437	0.0832	0.155
	0.6	0.1732	**0.0789**	0.1105	0.1885	0.1551	0.1488	0.1973	0.1493	0.0975	0.1627	0.0992	0.1628
	0.9	0.6575	0.5451	0.4853	**0.4382**	0.7311	0.8819	1.0841	0.8835	0.6199	0.7251	0.6557	0.6122
0.4	0.3	**0.0334**	0.0524	0.1088	0.1934	0.0368	0.0594	0.1454	0.0593	0.0739	0.1356	0.0728	0.1461
	0.6	0.166	**0.0803**	0.1116	0.1892	0.1565	0.1635	0.2385	0.1663	0.1	0.1714	0.1011	0.1639
	0.9	0.5619	0.5027	0.4597	**0.4249**	0.7508	1.1471	1.436	1.0075	0.679	0.8647	0.7115	0.6851

Table 5: Bias and mean square error (MSE) for all LPW-based estimators: $T = 500$. Unadjusted (LPW); analytically bias-adjusted (LPW-BA); bootstrap bias-adjusted (LPW$_{sb}$ for $k = 0, 1, 2$); bootstrap bias-adjusted *after* analytical adjustment (LPW-BA$_{sb}$). The lowest bias (in absolute value) and MSE for each parameter setting are highlighted in bold.

			$SPLW\text{-}BA$			$SPLW_{sb}$				$SPLW\text{-}BA_{sb}$			
										$P=1$			$P=2$
		$SPLW$	$P=1$	$P=2$	$P=3$	$k=0$	$k=1$	$k=2$	SSR	$k=0$	$k=1$	SSR	$k=0$
d	ϕ	Bias											
0	0.3	0.0573	-0.0058	-0.013	-0.032	0.0323	-0.0013	-0.0517	-0.0009	0.0012	0.0076	-0.0014	**0**
	0.6	0.2306	0.055	0.0068	-0.0255	0.1755	0.092	-0.0501	0.0922	0.0286	-0.0117	0.0293	**0.0005**
	0.9	0.725	0.5273	0.3849	**0.2659**	0.6762	0.6045	0.4876	0.605	0.4765	0.4002	0.477	0.334
0.2	0.3	0.0564	-0.0038	-0.0108	-0.0293	0.0316	-0.0018	-0.0513	-0.0012	0.003	0.0091	**0**	0.0009
	0.6	0.2292	0.0569	0.009	-0.0228	0.1716	0.0847	-0.063	0.0849	0.0292	-0.0122	0.0302	**0.0017**
	0.9	0.7195	0.5269	0.3854	**0.2679**	0.6846	0.6298	0.5374	0.6304	0.4696	0.3839	0.4685	0.3265
0.4	0.3	0.0582	0.0018	-0.0046	-0.0227	0.0316	-0.004	-0.0567	-0.0034	0.0048	0.007	0.0034	0.0024
	0.6	0.2296	0.0621	0.0146	-0.0163	0.1664	0.0719	-0.0889	0.0721	0.0292	-0.018	0.03	**0.0017**
	0.9	0.702	0.5222	0.3839	**0.2697**	0.7464	0.8296	0.8771	0.8312	0.4852	0.4267	0.4826	0.3283
d	ϕ	MSE											
0	0.3	**0.0075**	0.0106	0.0194	0.03	0.0076	0.0134	0.0356	0.0135	0.0137	0.0202	0.0133	0.0231
	0.6	0.0578	**0.0137**	0.0194	0.0295	0.0373	0.0241	0.0523	0.0241	0.018	0.0348	0.0181	0.0281
	0.9	0.5312	0.2907	0.1694	**0.101**	0.4652	0.3801	0.2747	0.3809	0.2461	0.1984	0.2477	0.1453
0.2	0.3	**0.0074**	0.0106	0.0196	0.0302	0.0075	0.0131	0.0345	0.0132	0.0134	0.0193	0.0129	0.0226
	0.6	0.0571	**0.0139**	0.0196	0.0298	0.0358	0.0223	0.0524	0.0224	0.0175	0.0332	0.0177	0.0276
	0.9	0.5232	0.2903	0.17	**0.1024**	0.4792	0.4201	0.3502	0.4209	0.2405	0.1878	0.2469	0.1403
0.4	0.3	0.0077	0.0108	0.0201	0.0305	**0.0075**	0.013	0.0344	0.0131	0.0131	0.0181	0.0128	0.0221
	0.6	0.0573	**0.0147**	0.0204	0.0302	0.0341	0.0205	0.057	0.0205	0.0173	0.0322	0.0175	0.0273
	0.9	0.4986	0.2854	0.1692	**0.1037**	0.5749	0.7557	1.1358	0.7496	0.2618	0.2452	0.2659	0.1461

Table 6: Empirical coverage and length of (nominal 95%) HPD intervals for all LPW-based estimators: $T = 100, 500$. Unadjusted (LPW); analytically bias-adjusted (LPW-BA); bootstrap bias-adjusted (LPW$_{sb}$ for $k = 0, 1, 2$); bootstrap bias-adjusted *after* analytical adjustment (LPW-BA$_{sb}$). Figures are averaged over all values of d used in the experimental design for each value of ϕ. Coverages for the intervals based on the asymptotic distribution of the LPW and analytically bias-adjusted (LPW-BA) estimators are also reported for comparison. The empirical coverage closest to the nominal 95%, and the shortest length, are highlighted in bold.

								$SPLW\text{-}BA_{sb}$			Asymptotic interval		
			$SPLW\text{-}BA$		$SPLW_{sb}$			$P=1$		$P=2$		$SPLW\text{-}BA$	
		$SPLW$	$P=1$	$P=2$	$k=0$	$k=1$	$k=2$	$k=0$	$k=1$	$k=0$	$SPLW$	$P=1$	$P=2$
ϕ	T	Coverage											
0.3	100	0.8563	0.9715	0.9643	0.8673	0.8773	0.858	0.9718	**0.963**	0.9663	0.6765	0.794	0.7565
	500	0.7883	0.9685	0.9658	0.8645	0.8938	0.901	0.959	**0.9448**	0.9648	0.789	0.8978	0.8575
0.6	100	0.14	0.9268	0.9663	0.1503	0.1768	0.2	0.9205	0.8955	**0.96**	0.0713	0.6768	0.748
	500	0.0438	0.93	0.9725	0.048	0.0613	0.0725	0.9375	**0.9385**	0.9663	0.0468	0.8505	0.864
0.9	100	0	0.1268	**0.5883**	0	0	0.0018	0.137	0.1549	0.5608	0	0.04	0.2205
	500	0	0.002	0.1018	0	0	0	0.0023	0.0045	**0.154**	0	0.0013	0.11
ϕ	T	Interval length											
0.3	100	**0.54**	0.9555	1.383	0.5413	0.5445	0.5529	0.9569	0.9587	1.3789	0.3911	0.5866	0.7332
	500	**0.263**	0.4289	0.577	0.2634	0.2645	0.2674	0.4291	0.4297	0.5775	0.2226	0.334	0.4175
0.6	100	**0.5434**	0.9562	1.3823	0.5445	0.548	0.5538	0.9586	0.9592	1.3756	0.3911	0.5866	0.7332
	500	**0.2676**	0.4305	0.5774	0.268	0.2712	0.2809	0.4313	0.434	0.5777	0.2226	0.334	0.4175
0.9	100	0.5034	0.8847	1.2921	0.4742	0.4248	**0.4073**	0.8236	0.79	1.1904	0.3911	0.5866	0.7332
	500	0.2638	0.4235	0.5753	0.269	0.2632	**0.2503**	0.427	0.4319	0.5785	0.2226	0.334	0.4175

References

Agiakloglou, C., Newbold, P. and Wohar, M. (1993). Bias in the estimator of the fractional difference parameter. *Journal of Time Series Analysis* **14** 235–246.

Andersen, T. G., Bollerslev, T., Christoffersen, P. F. and Diebold, F. X. (2006). Volatility and correlation forecasting. In G. Elliott, C. Granger and A. Timmermann, eds., *Handbook of Economic Forecasting*, no. 1 in Handbooks in Economics, chap. 15. Elsevier, 1st ed., 777–878.

Andrews, D. W. K. and Guggenberger, P. (2003). A bias-reduced log-periodogram regression estimator for the long-memory parameter. *Econometrica* **71** 675–712.

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*. 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.

Box, G. and Jenkins, G. (1970). *Time Series Analysis: Forecasting and Control*. Holden Day, San Francisco.

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.

Chibisov, D. M. (1972). An asymptotic expansion for the distribution of a statistic admitting an asymptotic expansion. *Theory of Probability and its Applications* **17** 620–630.

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.

Doukhan, P., Oppenheim, G. and Taqqu, M. S., eds. (2003). *Theory and Applications of Long-Range Dependence*. Birkhauser.

Durbin, J. (1960). The fitting of time series models. *Review of International Statistical Institute* **28** 233–244.

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.

Giraitis, L., Robinson, P. M. and Samarov, A. (1997). Rate optimal semiparametric estimation of the memory parameter of the gaussian time series with long-range dependence. *Journal of Time Series Analysis* **18** 49–60.

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.

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. (2002). Asymptotic behavior for partial autocorrelation functions of fractional ARIMA processes. *Annals of Applied Probability* **12** 1471–1491.

Inoue, A. and Kasahara, Y. (2004). Partial autocorrelation functions of the fractional ARIMA processes with negative degree of differencing. *Journal of Multivariate Analysis* **89** 135–147.

Künsch, H. R. (1989). The jacknife and the bootstrap for general stationary observations. *Annals of Statistics* **17** 1217–1241.

Levinson, N. (1947). The Wiener RMS (root mean square) error criterion in filter design and prediction. *Journal of Mathematical Physics* **25** 261–278.

Lieberman, O. (2001). The exact bias of the log-periodogram regression estimator. *Econometric Reviews* **20** 369–383.

Moulines, E. and Soulier, P. (1999). Broad band log-periodogram regression of time series with long range dependence. *Annals of Statistics* **27** 1415–1439.

Nielsen, M. . and Frederiksen, P. H. (2005). Finite sample comparison of parametric, semiparametric, and wavelet estimators of fractional integration. *Econometric Reviews* **24** 405–443.

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., Grose, S. D. and Martin, G. M. (2012). Higher order improvements of the sieve bootstrap for fractionally integrated processes. Econometrics & Business Statistics Working Paper WP09/12, 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.

Robinson, P. M. (2003). *Time Series with Long Memory*. Advanced Texts in Econometrics. Oxford University Press.

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.

Wand, M. and Jones, M. (1995). *Kernel Smoothing*. Chapman and Hall.