Econometrics and Statistics 29 (2024) 1–15

Contents lists available at ScienceDirect

Econometrics and Statistics

journal homepage: www.elsevier.com/locate/ecosta

Bootstrapping long memory time series: Application in low frequency estimators

Josu Arteche*

Dept. of Quantitative Methods, University of the Basque Country UPV/EHU, Bilbao 48015, Spain

ARTICLE INFO

Article history:
Received 14 December 2020
Revised 21 June 2021
Accepted 22 June 2021
Available online 8 July 2021

Keywords:
Long memory
Bootstrap
Memory parameter estimation

ABSTRACT

Bootstrapping time series requires dealing with the dependence that may exist within the sample. Several strategies have been proposed, but their validity has only been proven for short memory series and there has been little progress in their theoretical properties under long memory, where strong persistence may invalidate conventional techniques. The first contribution is to review all these recent advances, paying particular attention to those approaches that do not rely on parametric models and offering a guide for practitioners who wish to use them in semiparametric or nonparametric contexts. The second contribution is a Monte Carlo analysis of the applicability of these bootstrap techniques for approximating the distribution of low frequency estimators of the memory parameter based on spectral behaviour at frequencies close to the origin.

1. Introduction

Bootstrapping is currently one of the most popular tools for approximating the distributional characteristics of certain statistics whose distribution is unknown or difficult to handle. Originally proposed for independent observations, bootstrapping has been extended to cope with the serial dependence usually found in time series, although the strong dependence in long memory series sometimes requires specific techniques to form a valid resampling strategy. This paper reviews the advances made so far in the application of bootstrapping in long memory series, focusing particularly on those bootstraps that have been theoretically justified. A guide is thus offered for practitioners on how to bootstrap long memory series using techniques whose validity is supported by theory (with the concept of validity meaning here that the bootstrap distribution function converges to the asymptotic distribution function based on the original observations; see Cavaliere and Georgiev (2020) for alternative definitions). Additionally an extensive Monte Carlo study is included to analyse the performance of these techniques in approximating the distribution of some popular semiparametric estimators of the memory parameter.

Traditional bootstrap strategies proposed for time series belong to one of two categories: the first consists of resampling residuals obtained from an estimated parametric model as if they were iid. The validity of this strategy relies on the correct specification of the model to be estimated and on the consistency of the estimation method, conditions under which the residuals can be considered as approximately independent. The second category brings together non-parametric bootstrap

* Corresponding author.
E-mail address: josu.arteche@ehu.eus

https://doi.org/10.1016/j.ecosta.2021.06.002

strategies that avoid the need to impose parametric restrictions. Two different non-parametric bootstraps are traditionally used in this group: the sieve bootstrap is based on resampling the residuals obtained by fitting an autoregressive model of sufficiently large order that seeks to capture the first order dependence of the series; and the block bootstrap resamples approximately independent blocks of observations where the serial dependence is embedded in each block.

All these approaches have in common that the final resampling is implemented in approximately independent quantities, either residuals or blocks, which can be difficult to achieve in long memory series where the dependence remains significant between observations which are far apart. Bootstrapping long memory series requires proper consideration of that dependence to guarantee its validity. For example, the validity of the block bootstrap relies on two conditions: the blocks of observations must be approximately independent and the joint distribution of the observations inside the blocks must be approximately the same across blocks. Those conditions are satisfied in short memory time series, but the significant dependence between far-apart observations in long memory series may cause the block bootstrap to fail as the independence of the blocks in the bootstrap sample precludes mimicking the strong dependence in the underlying observations (see Lahiri, 1993 or Kim and Nordman, 2011). A natural alternative way to deal with this serial dependence is to apply prior filters to obtain approximately uncorrelated quantities for resampling. For example Andrews et al. (2006) propose resampling residuals obtained from the estimation of a stationary fractionally integrated ARMA (ARFIMA) model. Its validity relies on the correct specification of the ARFIMA to be estimated. To avoid the potential effects of misspecification, non-parametric methods have become popular tools in bootstrap strategies. In this context Poskitt (2008) shows the validity of the sieve bootstrap only for stationary and invertible ARFIMA models. More recently, a "de-colouring" approach has been proposed based on implementing the bootstrap not on the original observations y_t but rather on fractionally differenced series $(1-L)^{\hat{d}}y_t$ for $\hat{d}$ a consistent estimator of the memory parameter. The purpose of this filtering is to eliminate the strong persistence prior to applying a resampling strategy. The final bootstrap series are obtained by "re-colouring" the resampled series by fractional integration. This strategy has broadened the range of processes and bootstraps that can be used, including stationary and non-stationary long memory and many of the conventional bootstrap techniques that have been shown to be valid in short memory series (see Poskitt et al., 2015; Kapetanios and Papailias, 2011 and Kapetanios et al., 2019).

All the above approaches are developed in the time domain. Resampling methods in the frequency domain are also of special interest when dealing with time series as serial correlation of the observations is avoided, at least asymptotically, by using Fourier transforms. Franke and Härdle (1992) and Dahlhaus and Janas (1996) make use of this characteristic to propose a valid bootstrap strategy for short memory series based on resampling periodogram ordinates studentised with a consistent estimator of the spectral density. Estimation of the spectral density function is well documented in short memory time series, but it is much more difficult under long memory, where traditional techniques usually fail (see Arteche, 2015). Kim and Nordman (2013) focus on long memory series and suggest a parametric plug-in estimator of the spectral density to normalise the periodogram, showing that a resampling of the studentised periodogram can be used to consistently estimate the distribution of the parametric Whittle estimator as long as the model is fully and correctly specified. In a regression context, Hidalgo (2003) proposes a residual bootstrap in the frequency domain based on resampling discrete Fourier transforms of OLS residuals normalised with their modulus in order to approximate the distribution of the OLS estimator in linear regression models. He shows the validity of this strategy for stationary series, including long memory, but only in an OLS linear regression context.

Studentising the periodogram with an estimator of the spectral density function has the effect of generating asymptotically iid quantities, which permits resampling over the whole frequency band. A different approach which does not rely on any estimation of the spectral density, is to directly resample the periodogram but only in a neighbourhood of the frequency of interest. The rationale behind this approach is that the periodogram ordinates from dependent stationary series are not asymptotically identically distributed over the whole band of frequencies unless the series is white noise, as mean and variance are proportional to the spectral density function $f(\lambda)$. However, continuity of $f(\lambda)$ in short memory series leads Paparoditis and Politis (1999) to treat periodogram ordinates as if they were identically distributed but only in a shrinking band of frequencies. They suggest resampling periodogram ordinates locally in a neighbourhood of the frequency of interest. This strategy is valid for short memory series but it may fail with long memory at frequencies close to the spectral pole. To overcome this problem, Arteche (2020) proposes studentising the periodogram locally using an estimation of the form of the spectral density (apart from a constant) around the spectral pole. He shows that a local resampling scheme of the locally studentised periodogram is valid for approximating the distribution of the periodogram at frequencies close to the origin, which are the frequencies that contain relevant information on the persistence of the series.

The rest of the paper is organised as follows. Section 2 reviews the results to date on the application of bootstrapping in long memory time series. It starts by outlining the different definitions of long memory under which the validity of the bootstraps detailed in Section 2.2 and 2.3 has been proven. Section 2.2 focuses on bootstraps in the time domain and Section 2.3 describes resampling strategies in the frequency domain. Section 3 analyses the performance of these bootstrap strategies in approximating the distribution of low frequency estimators of the memory parameter by means of an extensive Monte Carlo study. Section 4 concludes.

2. Bootstrapping long memory: A review

2.1. Long memory

We consider linear processes of the form

$$y_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j},$$

where ε_t are iid variables with mean 0 and finite variance and a_j are real-valued constants. When $\sum_{j=0}^{\infty} a_j^2 < \infty$ the process is stationary. Long memory series are characterised by a significant dependence between far-apart observations such that the autocovariances satisfy

$$\gamma_y(k) = cov(y_t, y_{t-k}) \sim G k^{2d-1} \text{ as } k \to \infty, \tag{1}$$

where G is a finite constant and d is the memory parameter that governs the strong persistence of y_t. If $0 < d < 1/2$ the series is stationary but shows long memory such that the autocovariances are not summable. In that case the spectral density function (sdf) behaves as

$$f_y(\lambda) \sim C\lambda^{-2d} \text{ as } \lambda \to 0, \tag{2}$$

where C is a positive finite constant.

The concept of long memory is usually related to the concept of fractional integration although the two are not necessarily equivalent: y_t is integrated of order d (possibly fractional) denoted as $y_t \sim I(d)$ if $(1-L)^d y_t = u_t$, where L is the lag operator ($L^k y_t = y_{t-k}$) and $u_t \sim I(0)$, meaning in this case that u_t is weak dependent with finite, positive and continuous sdf over the whole frequency band $[0, \pi]$. If $y_t \sim I(d)$ then f_y satisfies (2). The most widely used $I(d)$ processes are the Fractionally Integrated ARMA (ARFIMA) parametric models, which assume that u_t is a stationary and invertible ARMA. The $ARFIMA(p, d, q)$ is defined as

$$\Phi_p(L)(1-L)^d y_t = \Theta_q(L)\varepsilon_t, \tag{3}$$

where $\Phi_p(L)$ and $\Theta_q(L)$ are polynomials of order p and q respectively with all roots outside the unit circle and no roots in common. In this case $y_t \sim I(d)$ with $u_t \sim ARMA(p, q)$.

In the case of non-stationary long memory ($d \geq 1/2$) the parameterisation and mathematical treatment of $I(d)$ processes are more complicated. The most notable effects are that $\sum_{j=0}^{\infty} a_j^2 = \infty$ such that the variance of the process is not finite and the sdf no longer exists, so the function in (2) can be considered as a pseudo-spectral density function. The modelling of non-stationary long memory series has moved in two different directions: Type I and Type II long memory (Robinson, 2005). They differ in their pre-sample treatments. Type I long memory processes are defined as

$$y_t = (1-L)^{-q} x_t \mathbb{I}(t \geq 1)$$

where $\mathbb{I}$ is the indicator function, q is a positive integer and x_t is a stationary $ARFIMA(p, b, q)$. With this definition $y_t \sim I(d)$ for $d = b + q$. Type II long memory processes are defined as

$$y_t = (1-L)^{-d} u_t \mathbb{I}(t \geq 1)$$

for $u_t \sim I(0)$, where it is assumed that the process is initialised at $y_0 = 0$. Note that Type II long memory processes are never stationary, but only asymptotically stationary when $d < 1/2$ whereas Type I processes are stationary if $q = 0$.

The theoretical contributions of different bootstrap strategies are provided based on different specifications of the long memory property; either parametric as in formula (3) or semiparametric as in (1) and (2). Their validity for approximating the distribution of statistics that are functions of observations of long memory processes has been analysed in a number of papers cited in Sections 2.2 and 2.3. In what follows, consider the long memory time series $Y_T = \{y_1, ..., y_T\}$ and let $Y_t^*\{y_1^*, ..., y_T^*\}$ be the corresponding bootstrap sample. Denote by $S_T = S(y_1, ..., y_T)$ the statistic of interest, which is a suitable smooth function of Y_T, and by $S_T^* = S(y_1^*, ..., y_T^*)$ the statistic evaluated at the bootstrap sample. Let $E(\cdot)$ and $E^*(\cdot)$ denote the expectations with respect to the pdf of Y_Y and Y_T^* respectively. Thus, S_T is usually required to satisfy the following condition:

Condition S: Let $\mathbb{G}$ be a compact subset of $\mathbb{R}^T$. Thus, for all $Y_T, Y_T^* \in \mathbb{G}$ there exists a family of Borel measurable functions $\mathbb{B}_t : \mathbb{R}^T \times \mathbb{R}^T \to [0, \infty)$ satisfying

$$\limsup_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E\left[E^*\left(\mathbb{B}_t(Y_T, Y_T^*)^2\right)\right] < \infty$$

for which

$$||S_T - S_T^*||^2 \leq \frac{1}{T^{1+\max\{0, d-1/2\}}} \sum_{t=1}^{T} \mathbb{B}_t(Y_T, Y_T^*)|y_t - y_T^*|^2.$$

This condition is satisfied by many common statistics such as the sample mean, autocovariances, autocorrelations, partial autocorrelations, discrete Fourier transforms, OLS regression coefficients and the Local Whittle estimator in the stationary case (see Poskitt, 2008 and Poskitt et al., 2015). However, the condition is considerably stronger when $d \geq 1/2$ and no longer allows for statistics such as the sample variance and sample autocorrelations, but still allows for the Local Whittle estimator (see Remark 15 in Kapetanios et al., 2019). Section 2.2 describes bootstrap methods for approximating the distribution of S_T in the time domain and analyses their validity with long memory series. Alternative strategies developed in the frequency domain are discussed in Section 2.3.

2.2. Bootstrap in the time domain

Andrews et al. (2006) were the first to show the theoretical advantages of bootstrap methods for long memory time series. They focus on stationary Gaussian time series with a spectral density function belonging to a parametric family satisfying

$$f_y(\lambda) = O\left(|\lambda|^{-2d-\delta}\right) \text{ as } \lambda \to 0, \ \forall \delta > 0 \text{ and } d \in (0, 1/2).$$

This specification is slightly weaker than the usual $f(\lambda) = O\left(|\lambda|^{-2d}\right)$ and is satisfied by the popular ARFIMA models in Equation (3). Once the model is estimated by maximum likelihood (ML) or by parametric Whittle estimation, a parametric bootstrap is proposed by simulating normal random variables with mean and covariance matrix obtained from the parametric model and the unknown parameters replaced by their estimated values. This parametric bootstrap provides higher-order improvements of the bootstrap confidence intervals over the coverages obtained with confidence intervals based on the asymptotic distribution of the ML or Whittle estimated parameters (similarly for linear tests). In particular the bootstrap CIs have smaller coverage probability errors than the CIs based on the asymptotic distribution by a multiplicative factor of order $o(T^{-1/2}\ln T)$. Despite this significant improvement, this bootstrap suffers from three main limitations: 1) The results depend on the model being specified correctly and completely; 2) Gaussianity is required; 3) Only the parametric bootstrap is considered, with other residual-based strategies being ignored. Cavaliere et al. (2015, 2017) extend the analysis and propose instead a wild bootstrap of the centred residuals in an ARFIMA model estimated by minimising the conditional sum-of-squares where Gaussianity is not required. The advantage of the wild bootstrap over other resampling approaches is that conditional and unconditional heteroscedasticity of a quite general and unknown form can be addressed in the implementation of tests for linear hypotheses on the long and short memory parameters of the model.

Nonparametric bootstrap methods that do not rely on a parametric model avoid the restrictions needed for the parametric bootstrap of Andrews et al. (2006): they do not need a full parametrisation of the model and Gaussianity is not an essential requirement. Two popular nonparametric strategies for time series, the Sieve bootstrap and the Block bootstrap, are described below and their applicability in long memory series is discussed. An additional strategy based on prior fractional differencing that is specifically designed for $I(d)$ processes is also discussed.

The Sieve Bootstrap (SBS)

The sieve bootstrap (SBS) is based on approximating the data-generating process by an autoregression of order h, AR(h), where h increases with the sample size. Bootstrap samples are then drawn from the residuals obtained from the autoregressive approximation. Poskitt (2008)) proposes the following procedure for stationary long memory series:

Step 1: Estimate the coefficient of an $AR(h)$ process fitted to the series Y_T. For practical purposes Burg's algorithm is suggested for the estimation of the coefficients and the selection of h is based on the minimisation of certain information criteria such as AIC. With the estimates of the AR coefficients $\hat{\phi}_1, \ldots, \hat{\phi}_h$ the residuals are obtained:

$$\hat{\varepsilon}_t = \sum_{j=0}^{h} \hat{\phi}_j y_{t-j}, \quad t = 1, \ldots, T,$$

with $y_{1-j} = y_{T-j+1}$, $j = 1, \ldots, h$, as initial values. Construct the standardised residuals

$$\hat{\varepsilon}_{s,t} = \frac{\hat{\varepsilon}_t - \bar{\varepsilon}}{S_{\hat{\varepsilon}}}, \quad \text{where } \bar{\varepsilon} = \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t \text{ and } S_{\hat{\varepsilon}} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\hat{\varepsilon}_t - \bar{\varepsilon}\right)^2}.$$

Step 2: Let ε_t^+, $t = 1, \ldots, T$ be a random sample of independent, identically distributed (iid) values drawn from a distribution placing a probability mass of $1/T$ at each $\hat{\varepsilon}_{s,t}$, $t = 1, \ldots, T$.

Step 3: Construct the bootstrap sample $Y_T^* = \{y_1^*, \ldots, y_T^*\}$ as

$$\sum_{j=0}^{h} \hat{\phi}_j y_{t-j}^* = \varepsilon_t^* =, \quad t = 1, \ldots, T,$$

with $y_{1-j}^* = y_{\tau-j+1}$, $j = 1, \ldots, h$, where τ has the discrete uniform distribution on the integers h,...,T, and $\varepsilon_t^* = \hat{\sigma}\varepsilon_t^+$ where $\hat{\sigma}^2$ is the estimated variance of the one-step-ahead prediction error obtained in Step 1 associated with the minimum mean square error predictor based on the infinite past.

Step 4: Obtain S_T^* as in S_T but with Y_T^* instead of Y_T.
Step 5: Repeat Steps 2, 3 and 4 B times to generate B independent bootstrap samples $Y_{T,b}^*$ and calculate $S_{T,b}^*$ for $b = 1, ..., B$. Approximate the distribution of S_T by the bootstrap distribution function

$$FS_{S_T^*,B}^*(s) = \frac{1}{B}\sum_{b=1}^{B}\mathbb{I}\{S_{T,b}^* \le s\}$$

where $\mathbb{I}$ is the indicator function.

Poskitt (2008) proves the validity of the SBS for linear stationary $I(d)$ processes with $|d| < 1/2$ and statistics S_T satisfying Condition S. The proof is based on the Mallows distance defined as follows: the Mallows distance between two probability distributions F_X and F_Y, denoted as $\eta(F_X, F_Y)$, is defined as $\eta(F_X, F_Y) = \inf\{E||X - Y||^2\}^{1/2}$, where the infimum is taken over all square integrable random variables X and Y with marginal distributions F_X and F_Y. Convergence in probability of the Mallows distance to zero implies convergence in distribution and convergence of the first two moments (see Lemma 8.3 9 in Bickel and Freedman, 1981). Poskitt (2008) shows that $\eta(FS_{S_T^*,B}^*, F_{S_T}) = o(1)$ a.s. as $T \to \infty$, and concludes that $FS_{S_T^*,B}^*(s)$ converges to the distribution of S_T with probability one, thus justifying the validity of the SBS in stationary long memory series.

The SBS strategy is based on a classical resampling of the standardized residuals in Step 2. This strategy is successful in mimicking the second order structure of the series, but may fail if higher order dependence also needs to be mirrored by the bootstrap samples. In that case, the wild bootstrap can be used in Step 2, as proposed by Fragkeskou and Paparoditis (2018) for weak dependent series. However, its validity for long memory series has not yet been theoretically justified.

The Block Bootstrap

The main alternatives to the sieve bootstrap for time series are based on resampling blocks of observations. The blocks can be disjoint, leading to the non-overlapping block bootstrap (NBB), or can overlap to form the moving block bootstrap (MBB). Let $l < T$ be an integer denoting the block length (either random or non-random) and let $\mathbb{B}_i = (y_i, ..., y_{i+l-1})$ denote a block with starting point $1 \le i \le T - l + 1$. A bootstrap series of length bl is defined as $(\mathbb{B}_{I_1}, ..., \mathbb{B}_{I_b})$ where $I_1, ..., I_b$ are iid uniform variables on $\{1, ..., T - l + 1\}$ for the MBB and on $\{1 + l(i - 1);\ i = 1, .., b\}$ for the NBB. Once the bootstrap series is obtained, the bootstrap statistic and its distribution are obtained as in Steps 4 and 5 of the SBS.

Approximate independence of the blocks is necessary for this resampling scheme to be successful. This may be far from being satisfied if the series have long memory. Lahiri (1993) was the first to analyse the applicability of the block bootstrap in long memory series. He focused on processes that satisfied the asymptotic behaviour requirement of the autocorrelations in Equation (1), showing that the MBB provides a valid approximation of the distribution of the normalized sample mean only if this statistic is asymptotically normal. In any other case, the MBB sample mean is asymptotically normal irrespective of the non-normal asymptotic distribution of the normalised sample mean. One reason for this failure is that the blocks are by construction independent in the bootstrap sample, which destroys the strong dependence that exists in the underlying observations. The effect is similar to the failure of Efron's bootstrap based on iid resampling schemes to capture the underlying dependence in the joint distribution of the observations in weak dependent processes, which makes it fail even in the simple case of the sample mean. Kim and Nordman (2011) extend Lahiri's results to show that the NBB is also valid in -not necessarily casual- linear long memory processes. In any case the conditional variance of the bootstrap mean has a slower growth rate than the variance of the average of the original observations (due to the aforementioned destruction of the strong dependence in the bootstrap samples), which entails a different normalisation for the bootstrap mean. In particular, for the bootstrap distribution to successfully mimic the distribution of the normalised sample mean the statistic must be multiplied by b^d, where b represents the number of blocks. In contrast with other traditional bootstraps, the empirical implementation of this normalisation requires a prior estimation of d. Moreover, this strategy cannot be generalised to statistics other than the sample mean and its validity has only been shown in stationary long memory series. These limitations can be overcome by the pre-filtering strategy approach proposed by Kapetanios and Papailias (2011) and discussed below. There are no more theoretical contributions on the application of the block bootstrap in long memory series. For their empirical performance see for example Butka and Beta (2014) and Butka and Puka (2014).

Pre-filtered bootstrap

A natural strategy for dealing with the problems raised by strong persistence is to eliminate it prior to implementing a bootstrap technique that is valid for weak dependent series. Poskitt et al. (2015) follows this idea and propose a modification of the SBS consisting of pre-filtering the series with the fractional difference operator constructed using a consistent estimator of the memory parameter prior to the application of the previous sieve bootstrap. With this modification the pre-filtered sieve bootstrap (PFSBS) consists of the following steps:

Step 0: Consider the fractional filter $(1 - L)^d = \sum_{j=0}^{\infty} a_j(d)L^j$ and let $\hat{d}$ be a consistent estimator of d. Generate the filtered series

$$\hat{u}_t = \sum_{j=0}^{t-1} a_j(\hat{d})y_{t-j}.$$

Steps 1, 2 and 3 of the sieve bootstrap applied to $\hat{u}_t$ instead of y_t. Obtain the sieve bootstrap sample $\hat{u}_t^*$.

Step 4: Obtain the pre-filtered sieve bootstrap sample

$$y_t^* = \sum_{j=0}^{t-1} a_j(-\hat{d})\hat{u}_{t-j}^*$$

and calculate S_T^* as in S_T but with Y_T^* instead of Y_T.

Step 5: Repeat Steps 1-4 B times to generate B independent bootstrap samples $Y_{T,b}^*$ and calculate $S_{T,b}^*$ for $b = 1, ..., B$. Approximate the distribution of S_T by the bootstrap distribution function

$$FPS_{S_T^*,B}^*(s) = \frac{1}{B}\sum_{b=1}^{B} \mathbb{I}\{S_{T,b}^* \le s\}.$$

Poskitt et al. (2015) restrict the analysis to the stationary case ($d < 1/2$) and show that the error of the PFSBS distribution is arbitrarily close to the traditional rate obtained with weak dependant series and improves the rate of the error of the SBS distribution as an approximation of the probability distribution of S_T for $d > 0$. In particular the error obtained with the SBS is $O_p(T^{d'-1+\beta})$ for any $\beta > 0$ and $d' = \max\{0, d\}$ whereas the error of the PFSBS is $O_p(T^{-1+\beta})$ if $\hat{d} - d = O_p(\log^{-1} T)$.

More recently Kapetanios et al. (2019) extend the results in Poskitt et al. (2015) by showing that the benefits of the prior fractional differencing in Step 0 remain valid with any bootstrap strategy whose properties are satisfied for weak dependent series, such as the block bootstrap or the spectral density-driven bootstrap of Krampe et al. (2018). An additional benefit of the pre-filtering strategy is that non-stationary long memory with $d \ge 1/2$ is also allowed by considering Type II fractionally integrated processes, in which case fractional differencing and integration in Steps 0 and 4 is complete, with no remaining terms. Kapetanios et al. (2019) show that the Mallows distance between the distribution of S_T satisfying Condition S and the bootstrap approximation based on pre-filtering goes to zero as the sample size increases as long as a consistent estimator of d exists, which proves the validity of pre-filtering combined with any bootstrap strategy valid with weak dependent series.

The accuracy of the pre-filtered bootstrap is analysed in Kapetanios et al. (2019) with a Monte Carlo analysis. It is concluded that a combination of pre-filtering and the spectral density-driven bootstrap of Krampe et al. (2018) is recommendable. After filtering as in Step 0 of the PFSBS, the pre-filtered spectral density-driven bootstrap (PFSDDB) is based on an estimation of the coefficients c_j of the Wold decomposition of $\hat{u}_t$ and consists of the following steps:

Step 1: Estimate consistently the spectral density function of $\hat{u}_t$ obtained as in Step 0 of the PFSBS, $\hat{f}_{\hat{u}}(\lambda)$, and compute the Fourier coefficients of $\log \hat{f}_{\hat{u}}(\lambda)$ as

$$\hat{a}_k = \frac{1}{2\pi}\int_0^{2\pi} \log(\hat{f}_{\hat{u}}(\lambda))e^{-ik\lambda}d\lambda$$

for $k = 0, 1, 2, ...$

Step 2: Compute $\hat{\sigma}^2 = 2\pi e^{\hat{a}_0}$ and the coefficients

$$\hat{c}_{k+1} = \sum_{j=0}^{k}\left(1 - \frac{j}{k+1}\right)\hat{a}_{k+1-j}\hat{c}_j$$

for $k = 0, 1, 2, ...$ and starting value $\hat{c}_0 = 1$.

Step 3: Generate iid pseudo innovations $\{\varepsilon_t^*, t \in \mathbb{Z}\}$ with mean 0 and variance $\hat{\sigma}^2$.

Step 4: Generate the bootstrap series $\hat{u}_t^*$, $t = 1, ..., T$ as

$$\hat{u}_t^* = \sum_{j=0}^{\infty} \hat{c}_j \varepsilon_{t-j}^*.$$

Step 5: Obtain the pre-filtered SDD bootstrap sample

$$y_t^* = \sum_{j=0}^{t-1} a_j(-\hat{d})\hat{u}_{t-j}^*$$

and calculate S_T^* as in S_T but with Y_T^* instead of Y_T.

Step 6: Repeat Steps 3, 4 and 5 B times to generate B independent bootstrap samples $Y_{T,b}^*$ and calculate $S_{T,b}^*$ for $b = 1, ..., B$. Approximate the distribution of S_T by the bootstrap distribution function

$$FPSDD_{S_T^*,B}^*(s) = \frac{1}{B}\sum_{b=1}^{B} \mathbb{I}\{S_{T,b}^* \le s\}.$$

In the above algorithm the bootstrap series $\hat{u}_t^*$ in Step 4 is generated using the estimated coefficients of the moving average representation. Alternatively, $\hat{u}_t^*$ can also be generated using the estimated coefficients of the autoregressive representation and modifying the algorithm appropriately, but no improvement has been shown with this modification (see Krampe et al., 2018 and Kapetanios et al., 2019).

2.3. Bootstrap in the frequency domain

Efron's iid bootstrap fails in dependent time series because the time dependence of the original observations is dismantled in the bootstrap samples. Different strategies such as the sieve or the block bootstrap seek to mimic the dependence of the original series in the bootstrap samples by finally resampling some functions of the series that are approximately independent, either residuals or blocks of observations. As an alternative, the frequency domain offers an advantageous context for the analysis of time series because the Fourier transform wipes out the serial dependence (at least asymptotically) turning autocorrelation into heteroscedasticity. The basic tool is the periodogram defined as

$$I_y(\lambda) = |W_y(\lambda)|^2 \ , \ \ W_y(\lambda) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} y_t \exp(-it\lambda)$$

where $W_y(\lambda)$ is the discrete Fourier transform (DFT) of y_t, $t = 1, .., T$ at frequency λ. Considering Fourier frequencies $\lambda_j = 2\pi j/T$, $j = 1, ..., \lfloor T/2 \rfloor$, where $\lfloor \rfloor$ denotes "the entire part of", one interesting property of $W_y(\lambda_j)$ is that they are asymptotically uncorrelated but possibly heteroscedastic, which avoids the problems caused by the existence of serial dependence. To avoid the heteroscedasticity of the raw DFT, Hidalgo (2003) proposes normalising the DFT and considers $W_y(\lambda_j)/|W_y(\lambda_j)|$, which can be regarded as a sequence of zero mean and asymptotically independent homoscedastic random variables regardless of the (weak) dependence of the original series. Hidalgo (2003) uses this characteristic and proposes resampling the normalised DFT of the OLS residuals in a simple linear regression model, showing that the distribution of the OLS estimator of the coefficient of the regression and the bootstrap approximation have the same limit distribution under stationary short and long memory in the disturbances and the regressor. However, only the OLS estimator is considered and no results on other interesting statistics are provided.

Many interesting statistics S_T can be written as functions of the periodogram. For example, the typical estimators of the autocovariances, of the spectral distribution function and even the parametric Whittle estimator can be expressed as weighted integrals of the periodogram. Estimators of this class are usually denoted as *spectral mean* estimators. Another important class of statistics is that of *ratio statistics*, which may be represented as ratios of spectral mean estimates and the integrated periodogram. For example, the sample autocorrelation is a ratio statistic. Bootstrapping in the frequency domain uses this characteristic and focuses on obtaining bootstrap samples of the periodogram, thus circumventing the need to generate bootstrap observations in the time domain, and therefore also avoiding the need to imitate the dependence structure of the underlying process. Franke and Härdle (1992) and Dahlhaus and Janas (1996) observe that periodogram ordinates of weak dependent series standardised by the spectral density are asymptotically iid and proposed a frequency domain bootstrap (FDB) strategy based on resampling periodogram ordinates studentised with a consistent estimator of the spectral density $I_y(\lambda_j)/\hat{f}_y(\lambda_j)$. Dahlhaus and Janas (1996) show the validity of this strategy for approximating the distribution of ratio statistics. However, it may fail when estimating certain spectral means (such as autocovariances) if the series is not Gaussian. In any case the validity of this strategy depends on the consistency of the estimation of the spectrum, which for short memory time series is simple and well documented but is especially difficult with long memory series, where traditional nonparametric techniques may fail (Arteche, 2015). This is noted by Kim and Nordman (2013), who propose using a plug-in parametric estimator of the spectral density function belonging to a parametric class of functions that includes long memory. The parameters are estimated using the parametric Whittle estimator and the estimates are plugged into the parametric form of the spectral density to obtain the Studentised periodogram, which is (after scaling by their mean) the quantity finally resampled to get the bootstrap approximation of the distribution of the Whittle estimator. Kim and Nordman (2013) show that this strategy leads to a consistent estimation of the distribution of the Whittle estimator if y_T is a stationary linear process with $d \in [0, 1/2)$.

The main disadvantage of the plug-in procedure in Kim and Nordman (2013) is that a complete parametric specification is required for the model on which the sdf is based. This limits its applicability for semiparametric estimators of d such as those considered in Section 5. These estimators, like other low frequency statistics, only need bootstrap replicates of the periodogram ordinates around the origin, which are the frequencies where the strong persistence plays the main role and where the standardised periodogram cannot be considered to be either independent or identically distributed. In fact, periodogram ordinates of long memory series are not asymptotically independent around the spectral pole and show a marked structure (far from the periodogram of a white noise) that needs to be replicated by the bootstrap samples. Taking this into account, Franco and Reisen (2004) and Silva et al. (2006) extend the proposal of Paparoditis and Politis (1999) to long memory series and suggest a local bootstrap strategy based on resampling periodogram ordinates only among neighbouring frequencies, thus retaining the global structure of the periodogram. Considering blocks of near periodogram ordinates makes this strategy resemble the more conventional block bootstrap. However, the block bootstrap is designed to maintain the local structure by resampling blocks while assuming independence between blocks but the local bootstrap resamples periodogram ordinates within a block of neighbouring frequencies, keeping the global structure of the periodogram unaltered. The benefits of this local bootstrap strategy have been justified only with simulations and no theoretical results have been offered unlike the results for weak dependent series in Paparoditis and Politis (1999). Moreover, as noted by Silva et al. (2006), this local bootstrap performs well only if a very narrow interval around the frequency of interest is used (in fact, Silva et al. (2006) propose resampling within a neighbourhood of only one or two frequencies), which induces problems in obtaining asymptotic results.

To overcome these problems and extend the number of frequencies subject to resampling, Arteche and Orbe (2016) propose that instead of the raw periodogram resampling should take place on the locally Studentised periodogram obtained by scaling the periodogram with a function that is proportional to the spectral density around the origin, namely $\lambda_j^{-2\hat{d}}$ for $\hat{d}$ a consistent estimator of d. This scaled periodogram shows a structure closer to the periodogram of a weak dependent series and the local bootstrap proposed by Paparoditis and Politis (1999) can retain the remaining structure of the locally studentised periodogram. The detailed procedure of this Frequency Domain Local Bootstrap (FDLB) is as follows:

Step 1: Estimate d consistently, say $\hat{d}$ and construct the locally studentised periodogram $\hat{v}_j = I(\lambda_j)\lambda_j^{2\hat{d}}$, for $j = 1, ..., \lfloor T/2 \rfloor$, where $\lfloor \cdot \rfloor$ denotes "the integer part of ".
Step 2: Select a resampling width $k_T \in \mathcal{N}$, $k_T \leq \lfloor T/2 \rfloor$.
Step 3: Define i.i.d. discrete random variables $J_1, ..., J_m$ taking values in the set $\Delta_T = \{0, \pm 1, ..., \pm k_T\} \backslash \{-j\}$ with probability p_i, $i \in \Delta_T$ (e.g. equal probability $p_i = 1/\#\Delta_T$ for all i).
Step 4: Generate B bootstrap series $\hat{v}^*_{bj} = \hat{v}_{|j+J_j|}$ for $b = 1, 2, ..., B$ and $j = 1, ..., m$.
Step 5: Generate B bootstrap samples for the periodogram $I^*_{bj} = \lambda_j^{-2\hat{d}}\hat{v}^*_{bj}$, for $b = 1, 2, ..., B$ and $j = 1, ..., m$.
Step 6: Obtain $S^*_{T,b}$ as S_T with I^*_{bj} replacing $I(\lambda_j)$ for $b = 1, 2, ..., B$.
Step 7: The bootstrap distribution of S_T is calculated as the empirical distribution of the B bootstrap replicates,

$$FLB^*_{S_{T,B}}(s) = \frac{1}{B}\sum_{b=1}^{B} \mathbb{I}\{S^*_{T,b} \leq s\}.$$

Note that this procedure resembles the bootstrap strategies in the time domain based on pre-filtering the observations with the fractional difference operator in that the goal of Step 1 is similar to the goal of pre-filtering as it aims to approximate the behaviour of $\hat{v}_j$ to the form of the periodogram of a weak dependent series where the local bootstrap has been proven to perform well. Although this FDLB strategy has been shown to perform very well in the approximation of the distribution of semiparametric estimators of d (Arteche and Orbe, 2016; 2017), its theoretical properties have only recently been analysed. Arteche (2020) considers linear long memory series with sdf satisfying Equation (2), also allowing for non-stationary long memory with $-1/2 < d < 1$. He offers a bound for the Mallows distance between the distribution of the periodogram and the bootstrap estimation using the previous FDLB strategy. This in turn is used to show that the Mallows distance $\eta(FLB^*_{S^*_T(m),B}, F_{S_T(m)}) = o(1)$ as $T \to \infty$ for $S_T(m) = \sum_{j=1}^{m} \psi_j I_y(\lambda_j)$ under some assumptions on the weights ψ_j and for m diverging to ∞ but more slowly than T. For example, these assumptions are satisfied in two relevant cases: 1) when $d < 1/2$ for $\psi_j = 2\pi/T$, in which case $S_T(m)$ is the average periodogram and estimates the spectral distribution at λ_m, which is used in the average-periodogram estimator of the memory parameter; 2) when $d < 3/4$ for $\psi_j = O(m^{-1/2}\lambda_j^{2d}\log m)$, with the validity of the FDLB being justified in this case for the score of the LW estimator in those situations where it is asymptotically normally distributed (the score of the LW estimator, and thus the LW estimator itself, are not asymptotically normal for $d \geq 3/4$, see Phillips and Shimotsu (2004) and Shao and Wu (2007).

3. Semiparametric estimation of *d*: A Monte Carlo analysis

One of the main applications of the different bootstrap strategies described above is to approximate the finite sample distribution of semiparametric estimators of d. Several papers propose the use of different bootstraps for that purpose, but most of them do not support their proposals with theoretical developments. The only theoretically supported strategies are those described above, at least to the author's knowledge. This section complements previous results by analysing their performance in approximating the distribution of three popular semiparametric estimators of the memory parameter: the log-periodogram estimator (LPE), the local Whittle estimator (LWE) and the average periodogram estimator (APE). The bootstraps considered are the SBS, PFSBS, PFSDDB, the MBB based on pre-filtering (PFMBB) and the FDLB.

3.1. Semiparametric or local estimation of d

Consider the following estimators of d:

Log-periodogram regression

The Log-Periodogram Estimator (LPE) $\hat{d}_{LP}$ is obtained by a simple least squares regression in the linear model

$$\log I_y(\lambda_j) = a + dX_j + u_j\,, \qquad j = 1, 2, ..., m, \tag{4}$$

where $X_j = -2\log\lambda_j$. The bandwidth m indicates the number of frequencies used in the estimation and is constrained to satisfy at least $m^{-1} + mT^{-1} \to 0$ as $T \to \infty$ such that even though $m \to \infty$ the band of frequencies used in the estimation degenerates to zero. The asymptotic properties of $\hat{d}$ are well established. In particular $\hat{d}$ is consistent for $-1/2 < d \leq 1$ and asymptotically normal for $-1/2 < d < 3/4$ such that

$$\sqrt{m}(\hat{d}_{LP} - d) \xrightarrow{d} N(0, \frac{\pi^2}{24}) \quad \text{as} \quad T \to \infty.$$

See Robinson (1995a); Hurvich et al. (1998); Velasco (1999a, 2000) and Kim and Phillips (2006). Arteche and Orbe (2005) propose a residual-based bootstrap and show that the bootstrap approximation of the LPE distribution is more accurate than the asymptotic distribution in small samples using an extensive Monte Carlo analysis. No theoretical justification has however been offered so far.

Local Whittle estimator

The Local Whittle Estimator (LWE) of the memory parameter d is obtained by minimising the function $R(d) = \log\left(m^{-1}\sum\lambda_j^{2d} I_y(\lambda_j) j\right) - m^{-1}2d\sum\log\lambda_j$. Its asymptotic properties are very similar to those of the LPE: if m satisfies at least $m^{-1} + mT^{-1} \to 0$ as $T \to \infty$ then $\hat{d}_{LW}$ is consistent for $-1/2 < d \le 1$ and asymptotically normal for $-1/2 < d < 3/4$

$$\sqrt{m}(\hat{d}_{LW} - d) \xrightarrow{d} N(0, \frac{1}{4}) \quad \text{as} \quad T \to \infty.$$

See Robinson (1995a) for the stationary case and Velasco (1999b); Phillips and Shimotsu (2004) and Shao and Wu (2007) for the nonstationary case. The validity of the bootstrap strategies considered in this section for approximating the distribution of the LWE is theoretically justified in Poskitt (2008); Poskitt et al. (2015); Kapetanios et al. (2019) and Arteche (2020). The finite sample performance of the FDLB is also analysed in Arteche and Orbe (2016), where its benefits are shown in terms of coverage of confidence intervals over the asymptotic distribution.

Average periodogram

The Average Periodogram Estimator (APE) was proposed by Robinson (1994) and is defined as

$$\hat{d}_{AP} = \frac{1}{2}\left\{1 - \frac{\log\left[\hat{F}(q\lambda_m)/\hat{F}(\lambda_m)\right]}{\log q}\right\} \text{ where } \hat{F}(\lambda) = \frac{2\pi}{T}\sum_{j=1}^{\lfloor\lambda T/2\pi\rfloor} I_y(\lambda_j)$$

for $q \in (0, 1)$ a user-chosen number. For practical purposes we use $q = 0.5$, as suggested by Lobato and Robinson (1996). Arteche (2020) provides some theoretical justification for the use of the FDLB to approximate the distribution of the APE in the stationary case $0 < d < 1/2$. Note however that the APE is asymptotically normal only for $0 < d < 1/4$, whereas the asymptotic distribution is non-normal for $d > 1/4$. In particular, Lobato and Robinson (1996) show that if $0 < d < 1/4$ then, as $T \to \infty$,

$$\sqrt{m}(\hat{d}_{AP} - d) \xrightarrow{d} N\left(0, \frac{(1 + q^{-1} - 2q^{-2d})}{\log^2 q}\frac{(1/2 - d)^2}{1 - 4d}\right), \tag{5}$$

but the asymptotic distribution has a non-normal intractable form for $1/4 < d < 1/2$. It is thus useful to analyse whether some of the bootstrap strategies discussed above, in particular the FDLB, can be successfully used to approximate the true distribution in all these situations, regardless of the tractability of the limiting distribution.

3.2. Comparison criteria

The performance of the asymptotic normal distribution and the different bootstrap strategies is analysed based on two different comparative measures: the root mean squared deviation of the bootstrap probability density function to the true density (obtained with Monte Carlo) and the coverage frequencies of confidence intervals obtained with the asymptotic distribution and the bootstrap estimates in different models. Both measures are obtained with R replications of series obtained from the DGP described in the next section.

The root mean squared deviation (RMSD)

The root mean squared deviation (RMSD) is defined as

$$RMSD(boot) = \sqrt{\frac{1}{R}\sum_{j=1}^{R}\left(p_{boot}(\hat{d}_j - d_0) - p_{mc}(\hat{d}_j - d_0)\right)^2}$$

where *boot* represents one of the bootstrap strategies discussed above, $p_{mc}(\hat{d}_j - d_0)$ is the ordinate of a kernel based density estimate obtained with the R estimates $\hat{d}_k - d_0$, $k = 1, \ldots, R$ evaluated at $\hat{d}_j - d_0$ and $p_{boot}(\hat{d}_j - d_0)$ is the average over the R simulations of the ordinates of kernel density estimates obtained using the B bootstrap estimates $\hat{d}_j^b - \hat{d}_j$, $b = 1, \ldots, B$ evaluated at $\hat{d}_j - d_0$ (see Poskitt et al., 2015). The closer this value is to zero the better the performance of the bootstrap considered is.

Coverages of confidence intervals (CI)

The CIs based on the asymptotic distribution for a confidence level $1 - \alpha$ are of the form

$$CI_{1-\alpha}(\hat{d}_a) = \left(\hat{d}_a + \sqrt{\widehat{var}(\hat{d}_a)}z_{\alpha/2}; \hat{d}_a + \sqrt{\widehat{var}(\hat{d}_a)}z_{1-\alpha/2}\right)$$

for $a = LW, LP, AP$, where z_α is the 100αth% percentile of a standard normal distribution $N(0, 1)$. The variance of the ALPE is obtained by plugging the estimate of d into the variance of the distribution in (5). The variance of the LWE is estimated as

$$\widehat{var}(\hat{d}_{LWE}) = \left(4\sum_{j=1}^{m}\left[\log\lambda_j - \frac{1}{m}\sum_{k=1}^{m}\log\lambda_k\right]^2\right)^{-1},$$

as suggested by Hurvich and Chen (2000) and Arteche (2006), which gives much better results in terms of coverage of CIs than the variance in the asymptotic distribution (see also Arteche and Orbe, 2016. For the same reason the OLS variance is used as $\widehat{var}(\hat{d}_{LPE})$ instead of the variance in the asymptotic distribution (see Arteche and Orbe, 2005).

The coverages of CIs based on the asymptotic distribution are compared with the CIs obtained with the different bootstrap strategies. We construct percentile CIs of the form

$$CI^*_{1-\alpha}(\hat{d}_a) = \left(\hat{d}^*_{a|\alpha/2}; \hat{d}^*_{a|1-\alpha/2}\right)$$

where $\hat{d}^*_{a|\alpha}$ denotes the $B\alpha$ ordered value of the bootstrap estimates of d for $a = LW, LP$ or AP. Arteche and Orbe (2005) considered alternative techniques, such as the percentile-t or the BCa when resampling the residuals of the LPE, and obtain that the percentile-t performs better than the others. Note however that the percentile$-t$ only makes sense with the LPE because its variance changes with the bootstrap sample, whereas the variance of the LWE, $\widehat{var}(\hat{d}_{LWE})$ is invariant to the sample. Regarding the APE, an expression for the variance is only available for $0 < d < 1/4$ and even in that case the use of the percentile$-t$ should be avoided because the variance strongly depends on d (see MacKinnon (2006). As a consequence only the LPE might benefit from using the percentile-t. However, symmetric and asymmetric versions of the percentile-t CIs have also been analysed and no improvement has been obtained over the percentile CI, because they have shown larger intervals without improving the coverage (results available on request).

We consider coverage discrepancy plots as suggested by de de Peretti and Siani (2010) defined as follows. For a nominal coverage x the coverage discrepancy is calculated as $\hat{F}(x) - x$, where $\hat{F}(x) = R^{-1}\sum_{r=1}^{R}\mathbb{I}(x_s \le x)$ for $x_s = 1 - 2\min(pv, 1 - pv)$ and pv is the estimated probability distribution function at the true memory parameter d_0, either using the asymptotic distribution or any of the bootstrap strategies described above. $\hat{F}(x)$ can thus be interpreted as the frequency of coverages of the true d_0 with CIs obtained with the estimated distribution used to get pv for a nominal confidence x. The closer the discrepancy is to zero the better the strategy used to approximate the distribution of the estimator of d_0 performs.

We pay particular attention to the 95% CI, which is frequently used for interval estimation and inference. Not only is the frequency of coverages important to assess the performance of a CI but its length is also a relevant measure of its effectiveness (see de Peretti and Siani, 2010). Greater coverage is usually associated with wider intervals but in some cases narrower intervals are accompanied by greater coverages, indicating higher precision in the estimation. That is, if two CIs have the same confidence level the shorter is preferable.

3.3. Monte Carlo design

The applicability of the previous bootstrap strategies is analysed by comparing their performance in $R = 1000$ replications of different ARFIMA models of the form

$$(1 - \phi L)(1 - L)^d X_t = u_t, \quad t = 1, 2, ..T, \tag{6}$$

where the u_t are standard independent normal. Two different sample sizes ($T = 64$ and $T = 128$) and three different bandwidths ($m = 3, 8, 15$ for $T = 64$ and $m = 5, 15, 30$ for $T = 128$) are considered. Two different values of the autoregressive parameter ($\phi = 0$ and $\phi = 0.6$) and four values of the memory parameter ($d = -0.4, 0.2, 0.4$ and 0.7) are analysed. The LWE and the LPE are consistent and asymptotically normal for all four values of d, but the limiting distribution of the APE is only normal for $d = 0.2$. For $d = 0.4$ it has an intractable limiting distribution. We include the values $d = -0.4$ and $d = 0.4$ in the analysis to establish whether the different bootstrap strategies can improve a perhaps wrongly assumed limiting normal distribution. The APE for the non-stationary value $d = 0.7$ is inconsistent and is thus not included in the analysis.

All the pre-filtered bootstrap strategies are based on a preliminary estimation of d obtained by LWE with an initial bandwidth m_1. The order of the autorregression in the SBS and PFSBS is selected by minimising the AIC and the coefficients estimated by Burg's algorithm. Three different block lengths are used for the MBB: $b_1 = \lfloor T^{1/5} \rfloor$, $b_2 = \lfloor T^{1/3} \rfloor$ and the optimal data-dependent block length proposed in Patton et al. (2009), denoted as $b*$. Finally, two different resampling widths are considered for the FDLB: $(k_1, k_2) = (5, 20)$ for $T = 64$ and $(k_1, k_2) = (10, 40)$ for $T = 128$. The number of bootstrap replications is 1000.

3.4. Results

For the sake of brevity only the representative results for $d = 0.2$ are included in this section, for a sample size of $T = 128$. In this case the three estimators considered are consistent and asymptotically normal, and thus the use of the normal distribution is theoretically justified. The results for the rest of the models and sample sizes are available upon request but the main conclusions are qualitatively similar.

Table 1
Ratio of RMSD for ARFIMA(1,0.2,0), $T = 128$.

	$m = 5$			$m = 15$			$m = 30$		
	$m_1 = 5$	$m_1 = 15$	$m_1 = 30$	$m_1 = 5$	$m_1 = 15$	$m_1 = 30$	$m_1 = 5$	$m_1 = 15$	$m_1 = 30$
	LWE, $\phi = 0$								
PFSBS	0.413	0.781	0.816	1.179	0.217	0.596	1.457	0.734	0.272
PFSDDB	0.366	0.790	0.836	1.577	**0.207**	0.621	2.278	1.248	**0.165**
PFMBB(b_1)	0.416	0.746	0.799	1.738	0.414	0.721	2.637	1.029	0.437
PFMBB(b_2)	0.369	0.746	0.802	1.425	0.358	0.641	1.838	0.640	0.361
PFMBB($b*$)	**0.365**	0.745	0.802	1.439	0.405	0.703	1.980	1.083	0.421
FDLB(k_1)	0.560	0.598	0.682	1.021	0.443	0.377	0.911	0.234	0.295
FDLB(k_2)	0.451	0.754	0.780	1.903	0.261	0.526	3.083	1.338	0.188
	LPE, $\phi = 0$								
PFSBS	0.445	0.776	0.811	1.138	0.401	0.698	1.136	0.741	0.536
PFSDDB	0.391	0.781	0.828	1.461	0.329	0.718	1.716	1.025	0.502
PFMBB(b_1)	0.263	0.719	0.769	1.565	0.422	0.796	2.029	0.862	0.494
PFMBB(b_1)	**0.212**	0.719	0.774	1.301	0.435	0.736	1.424	0.666	0.510
PFMBB($b*$)	0.272	0.721	0.773	1.347	0.444	0.777	1.505	0.908	0.479
FDLB(k_1)	0.533	0.641	0.709	0.897	**0.223**	0.429	0.670	0.291	**0.258**
FDLB(k_2)	0.479	0.746	0.777	1.801	0.245	0.622	2.467	1.106	0.443
	APE, $\phi = 0$								
PFSBS	0.722	0.885	0.895	1.797	1.058	1.127	2.390	1.689	1.121
PFSDDB	0.721	0.894	0.916	2.114	1.067	1.148	3.349	2.470	1.046
PFMBB(b_1)	0.810	0.879	0.894	2.408	1.298	1.309	3.804	2.264	1.362
PFMBB(b_2)	0.778	0.878	0.896	2.036	1.232	1.233	2.937	1.636	1.278
PFMBB($b*$)	0.785	0.877	0.894	1.983	1.302	1.302	2.941	2.263	1.364
FDLB(k_1)	**0.622**	0.724	0.779	1.644	**0.974**	0.983	1.974	**0.909**	0.972
FDLB(k_2)	0.743	0.869	0.872	2.468	1.045	1.080	4.458	2.681	0.955
	LWE, $\phi = 0.6$								
PFSBS	0.168	0.983	1.322	0.807	0.718	0.421	0.959	0.996	0.976
PFSDDB	**0.091**	0.997	1.402	0.923	0.913	0.348	**0.913**	1.011	1.011
PFMBB(b_1)	0.099	1.008	1.422	1.088	0.926	**0.329**	0.924	1.012	1.007
PFMBB(b_2)	0.122	1.012	1.408	1.021	0.887	0.349	0.945	1.020	1.011
PFMBB($b*$)	0.087	1.012	1.405	1.009	0.875	**0.329**	0.927	1.011	1.011
FDLB(k_1)	0.357	0.785	1.226	0.987	0.918	0.391	0.962	1.023	1.008
FDLB(k_2)	0.802	1.066	1.341	1.086	0.974	0.376	0.945	1.011	1.016
	LPE, $\phi = 0.6$								
PFSBS	0.376	0.835	1.027	0.862	0.696	0.514	0.954	0.951	**0.874**
PFSDDB	0.381	0.840	1.086	0.928	0.808	0.451	0.921	0.993	0.926
PFMBB(b_1)	0.354	0.841	1.093	1.134	0.809	**0.427**	0.945	0.996	0.905
PFMBB(b_2)	0.360	0.846	1.084	1.080	0.790	0.454	0.959	0.995	0.905
PFMBB($b*$)	**0.344**	0.844	1.084	1.036	0.778	0.434	0.946	0.991	0.919
FDLB(k_1)	0.458	0.621	0.898	0.997	0.790	0.471	0.999	1.011	0.954
FDLB(k_2)	0.776	0.887	1.023	1.137	0.890	0.459	0.966	1.003	0.942
	APE, $\phi = 0.6$								
PFSBS	0.973	1.155	1.385	0.759	0.683	0.367	0.994	0.998	1.000
PFSDDB	0.943	1.160	1.428	0.853	0.836	0.327	0.977	0.995	1.000
PFMBB(b_1)	0.948	1.170	1.444	1.100	0.833	0.312	0.973	0.995	1.000
PFMBB(b_2)	**0.932**	1.170	1.433	1.027	0.804	0.349	0.980	0.998	1.002
PFMBB($b*$)	0.956	1.175	1.430	0.970	0.797	**0.304**	0.979	0.995	1.001
FDLB(k_1)	0.947	1.124	1.391	0.983	0.832	0.389	0.987	0.999	1.001
FDLB(k_2)	1.136	1.215	1.406	1.085	0.933	0.381	**0.969**	0.994	0.999

Note: The numbers in each cell show the ratio of RMSD obtained with the different bootstrap strategies with respect to the sieve. $b_1 = \lfloor T^{1/5} \rfloor$, $b_2 = \lfloor T^{1/3} \rfloor$, $b*$ is the optimal data-dependent block length of Patton et al. (2009) and $(k_1, k_2) = (10, 40)$.

Table 1 shows ratios of the RMSD of the different bootstrap strategies over RMSD(SBS). In general the ratio is lower than one, indicating that pre-filtering either in the time domain or in the frequency domain leads to bootstrap distributions that are closer to the true distribution than the approximation offered by the SBS. The main disadvantage of pre-filtering is that user intervention is required for the selection of m_1. As a rule of thumb, a value $m_1 = m$ is a wise choice, especially when $\phi = 0$, but when $\phi = 0.6$ the bias in the estimation of d implied by the use of a large bandwidth means that a lower m_1 may lead to better results by avoiding the bias in the estimation of d used to pre-filter. No definitive suggestion can be made as regards the selection of k_T for the FDLB. Values of m_1 and k_T close to m leads to general good results and a large k_T is significantly more harmful than a lower one when $m_1 < m$, especially when there is no short memory component. Less information can be extracted for the selection of the block length in the PFMBB. The "optimal" $b*$ is not uniformly superior to the other two lengths considered, which sometimes and with no prevalence may offer superior results.

Figure 1 shows the coverage discrepancies in the illustrative case of an $ARFIMA(1, 0.2, 0)$ with $T = 128$ and $m = m_1 = 15$. For the sake of clarity of exposition and visibility only the results obtained with the asymptotic distribution and the boot-

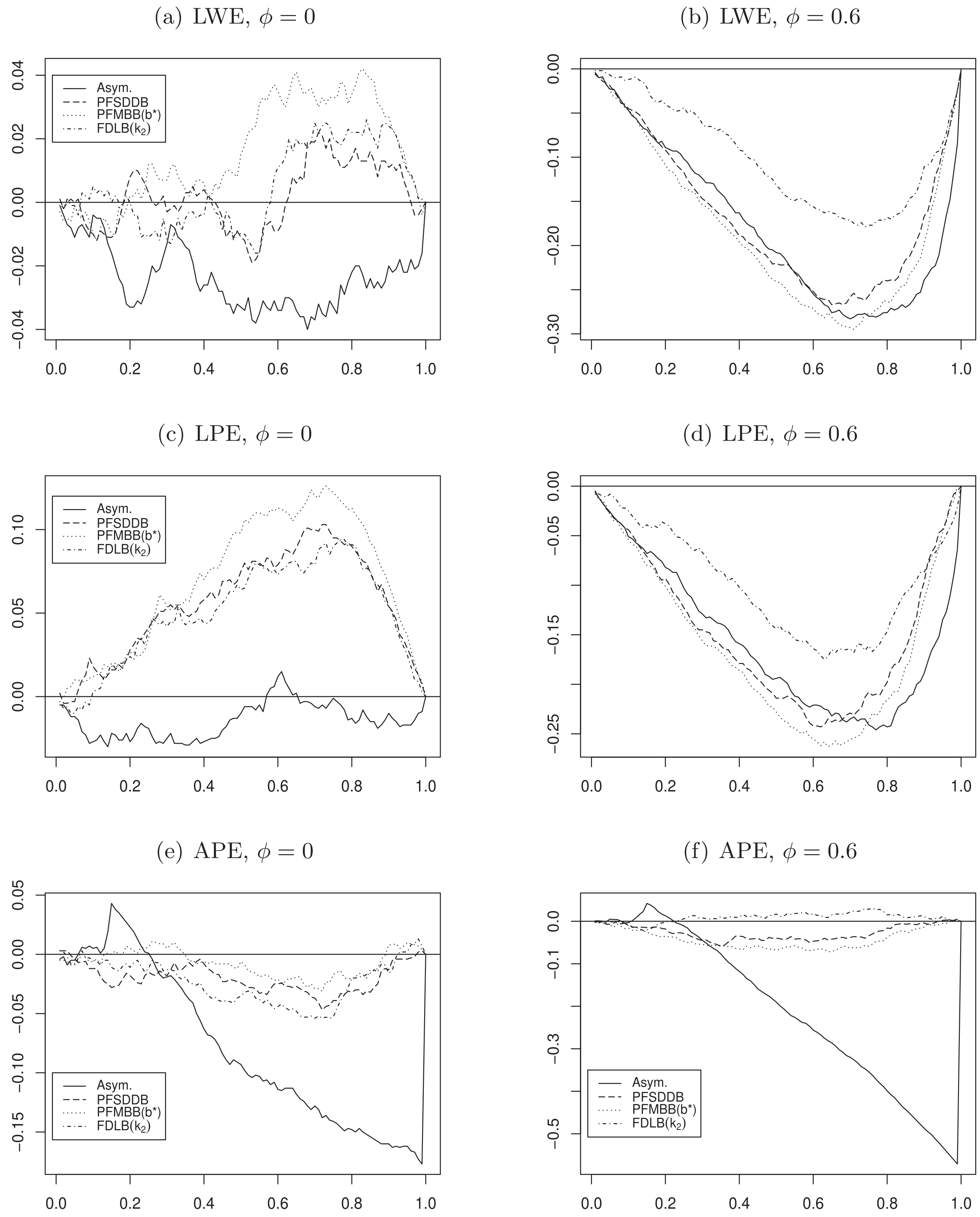

Fig. 1. Coverage discrepancies for ARFIMA(1,0.2,0), $T = 128$, $m = m_1 = 15$

strap strategies PFSDDB, PFMBB(b^*) and FDLB (k_2) are included. This set of resampling techniques suffices to illustrate the advantages of bootstrapping over the asymptotic distribution. Bootstrapping tends to offer lower discrepancies except in the LPE case when $\phi = 0$, where the coverage with the asymptotic distribution tends to be better. Note that no theoretical results exist as to the validity of these bootstraps with the LPE, although Arteche and Orbe (2005) show the benefits of alternative residual resampling strategies not considered here. The poor performance of the asymptotic distribution of the APE is also noteworthy, even when $d = 0.2$, which belongs to the interval of memory parameters where asymptotic normality has been rigorously proved. In this case any of the bootstraps considered significantly improves coverage. When $\phi = 0.6$ the

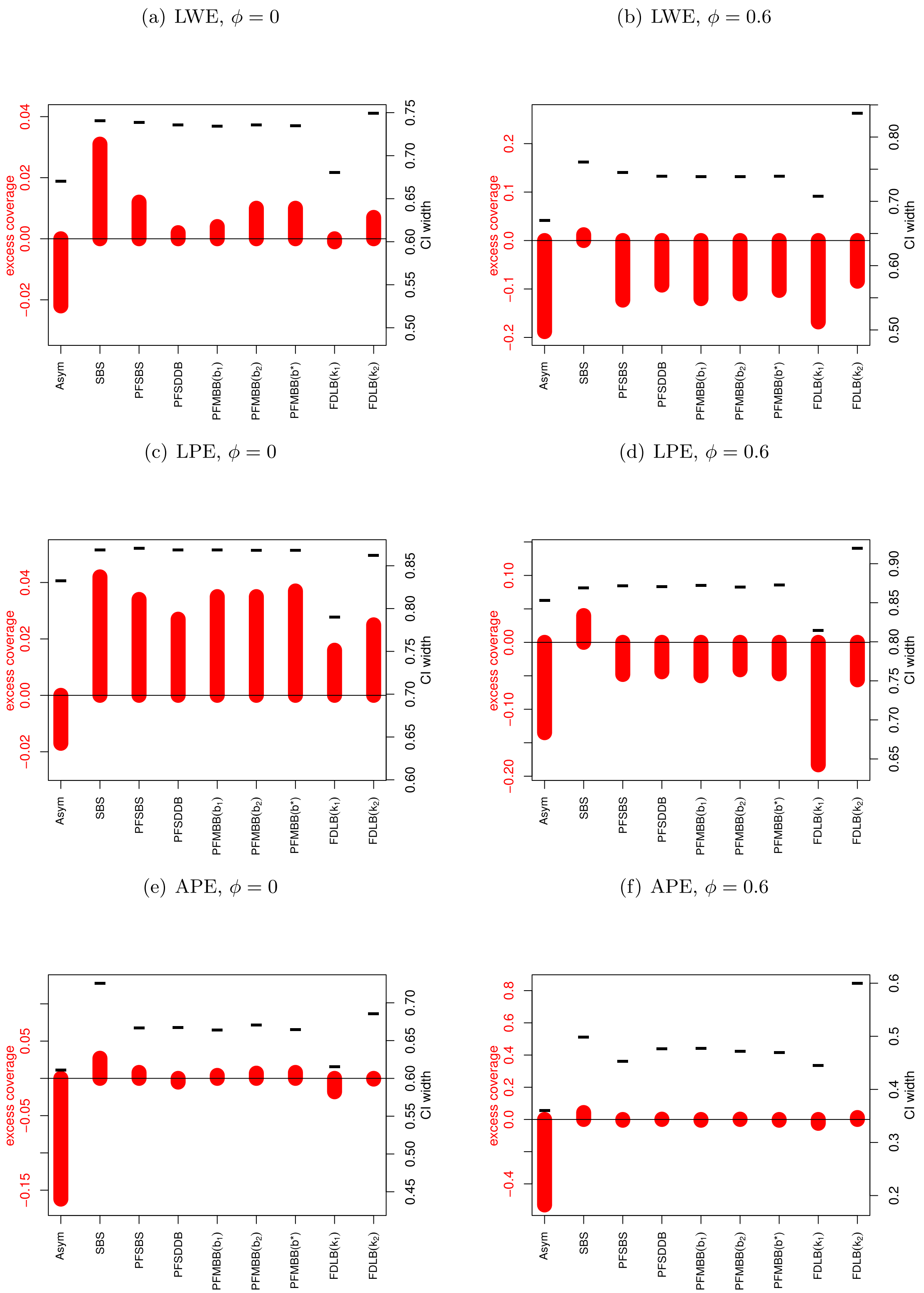

Fig. 2. 95% coverage discrepancies and CI lengths for ARFIMA(1,0.2,0), $T = 128$, $m = m_1 = 15$ *Note:* The red bars show the excess of coverage (in red left vertical axis) and the horizontal lines show the average lengths (right vertical axis).

bias in the estimation of the memory parameter naturally arises, significantly affecting the coverage, but the performance of the FDLB (k_2) is more robust to this biasing effect.

Finally, Figure 2 focuses on the coverage of 95% CIs. The vertical bars show the excess of coverage (actual coverage minus 0.95) with measures on the left vertical axis (in red) and the horizontal lines show the average lengths (over the 1000 simulations) of the CIs (right vertical axis). The asymptotic distribution undercovers, with actual coverages below the nominal 0.95, but some of the bootstrap alternatives give closer-to-zero discrepancies even with shorter intervals in some cases, for example the FDLB(k_1) in the LPE and the APE with $\phi = 0$.

The results obtained in the rest of the models, sample sizes and bandwidths can be summarised as follows. When $\phi = 0$ the asymptotic CIs of the LWE and the LPE (with large m) give acceptable coverages but the undercoverage of the APE is more significant, especially when $d = 0.4$ with a medium-large m. Note that the APE is not asymptotically normal in this case, so the CIs obtained under the normal distribution are unfounded. Including the short memory component generates a bias in the estimation that affects the coverages of all the different CIs. However, bootstrapping can significantly improve coverage in those cases where the bias is significant (large m) and the asymptotic CIs clearly undercover. In general, there is no strategy that performs uniformly better than the others, but it is always possible to find a bootstrap leading to CIs with coverages closer to the nominal confidence level. As commented above for the RMSD, pre-filtering is a beneficial approach and $m_1 = m$ seems a good option. Furthermore, the SBS (for intermediate m) and the FDLB (for large m) are in general the strategies that lead to the coverages least affected by the bias induced by the short memory component.

4. Conclusion

Bootstrapping has become a popular technique for approximating the distribution of many statistics of time series. Its use with long memory has increased significantly in the last few years although there exist few theoretical justifications of its validity in the context of strong persistence. This paper reviews recent results on the theoretical properties of some bootstrap strategies applied to long memory time series, with particular attention on their use in approximating the distribution of low frequency estimators of the memory parameter. All the techniques described in this paper share the characteristic of not requiring full parametric restrictions, which makes them suitable for application with non-parametric and semi-parametric estimation techniques.

A natural, successful strategy for dealing with strong persistence consists of *decolouring* the series before applying bootstrap techniques whose validity has been proven with short memory series. The main drawback of this approach is that a prior estimation of the memory parameter is required, which in turn entails selecting a prior bandwidth m_1 and makes the whole procedure non-automatic, unlike other techniques not based on prior *decolouring* such as the sieve bootstrap. However, simulations suggest that a bandwidth $m_1 = m$ is a sensible choice, leading to much better results for the prefiltered SBS or SDDB than the sieve bootstrap in many situations. Other approaches, such as the block bootstrap or the FDLB, also require selection of the block lengths or the resampling width for the FDLB. The simulations suggest that existing techniques for optimal selection of the block length such as the proposal by Patton et al. (2009) do not necessarily give better results than other naive selections. Some ad hoc criteria for selecting the resampling width in the FDLB are proposed by Arteche and Orbe (2016), but there is no automatic criterion and further research on this topic and on the selection of the block length in block bootstrapping in the context of long memory series is needed to extend their use among empirical researchers.

Acknowledgments

Research supported by the Spanish Ministry of Science and Innovation and ERDF grants ECO2016-76884-P, PID2019-105183GB-I00 and UPV/EHU Econometrics Research Group (Basque Government grant IT1359-19). I thank the Associate Editor and two referees for helpful and constructive comments that led to significant improvements in the contents and presentation of the paper.

References

Andrews, D.W.K., Lieberman, O., Marmer, V., 2006. Higher-order improvements of the parametric bootstrap for long-memory time series. Journal of Econometrics 133, 673–702.
Arteche, J., 2006. Semiparametric estimation in perturbed long memory series. Computational Statistics and Data Analysis 51, 2118–2141.
Arteche, J., 2015. Signal extraction in long memory stochastic volatility. Econometric Theory 31, 1382–1402.
Arteche, J., 2020. Frequency domain local bootstrap in long memory time series. BILTOKI Working Paper. http://hdl.handle.net/10810/48980.
Arteche, J., Orbe, J., 2005. Bootstrapping the log-periodogram regression. Economics Letters 86, 79–85.
Arteche, J., Orbe, J., 2016. A bootstrap approximation for the distribution of the local whittle estimator. Computational Statistics and Data Analysis 100, 645–660.
Arteche, J., Orbe, J., 2017. A strategy for optimal bandwidth selection in local whittle estimation. Econometrics and Statistics 4, 3–17.
Bickel, P.J., Freedman, D.A., 1981. Some asymptotic theory for the bootstrap. Annals of Statistics 9, 1196–1217.
Butka, A., Beta, D., 2014. Bootstrap estimation of long-range dependence in arfima processes. Journal of Advances in Mathematics 7, 1267–1276.
Butka, A., Puka, L., 2014. A block bootstrap procedure for long memory processes. International Journal of Mathematics Trends and Technology 14, 72–78.
Cavaliere, G., Georgiev, L., 2020. Inference under random limit bootstrap measures. Econometrica 88, 2547–2574.
Cavaliere, G., Nielsen, M.O., Taylor, A.M.R., 2015. Bootstrap score tests for fractional integration in heteroskedastic ARFIMA models, with an application to price dynamics in commodity spot and futures markets. Journal of Econometrics 187, 557–579.
Cavaliere, G., Nielsen, M.O., Taylor, A.M.R., 2017. Quasi-maximum likelihood estimation and bootstrap inference in fractional time series models with heteroskedasticity of unknown form. Journal of Econometrics 165–188.

Dahlhaus, R., Janas, D., 1996. A frequency domain bootstrap for ratio statistics in time series analysis. Annals of Statistics 24, 1934–1963.
Fragkeskou, M., Paparoditis, E., 2018. Extending the range of validity of the autoregressive (sieve) bootstrap. Journal of Time Series Analysis 39, 356–379.
Franco, G.C., Reisen, V.A., 2004. Bootstrap techniques in semiparametric estimation methods for ARFIMA models: A comparison study. Computational Statistics 19, 243–259.
Franke, J., Härdle, W., 1992. On bootstrapping kernel spectral estimates. Annals of Statistics 20, 121–145.
Hidalgo, J., 2003. An alternative bootstrap to moving blocks for time series regression models. Journal of Econometrics 117, 369–399.
Hurvich, C.M., Chen, W.W., 2000. An efficient taper for potentially overdifferenced long-memory time series. Journal of Time Series Analysis 21, 155–180.
Hurvich, C.M., Deo, R., Brodsky, J., 1998. The mean squared error of geweke and porter-hudak's estimator of the memory parameter in a long-memory time series. Journal of Time Series Analysis 19, 19–46.
Kapetanios, G., Papailias, F., 2011. Block bootstrap and long memory. working paper 679. School of Economics and Finance, Queen Mary, University of London.
Kapetanios, G., Papailias, F., Taylor, A.M.R., 2019. A generalised fractional differencing bootstrap for long memory processes. Journal of Time Series Analysis 40, 467–492.
Kim, C.S., Phillips, P.C.B., 2006. Log periodogram regression: The nonstationary case. Cowles Foundation Discussion Paper No.1587. Available at SSRN: https://ssrn.com/abstract=937915.
Kim, Y.M., Nordman, D.J., 2011. Properties of a block bootstrap under long-range dependence. Sankhya Ser A 73, 79–109.
Kim, Y.M., Nordman, D.J., 2013. A frequency domain bootstrap for whittle estimation under long-range dependence. Journal of Multivariate Analysis 115, 405–420.
Krampe, J., Kreiss, J.P., Paparoditis, E., 2018. Estimated wold representation and spectral-density-driven bootstrap for time series. Journal of the Royal Statistical Society: Series B 80, 703–726.
Lahiri, S.N., 1993. On the moving block bootstrap under long range dependence. Statistics & Probability Letters 18, 405–413.
Lobato, I., Robinson, P.M., 1996. Averaged periodogram estimation of long memory. Journal of Econometrics 73, 303–324.
MacKinnon, J.G., 2006. Bootstrap methods in econometrics. The Economic Record 82, S2–S18.
Paparoditis, E., Politis, D., 1999. The local bootstrap for periodogram statistics. Journal of Time Series Analysis 20, 193–222.
Patton, A., Politis, D., White, H., 2009. Correction to "automatic block-length selection for the dependent bootstrap". Econometric Reviews 23, 53–70.
de Peretti, C., Siani, C., 2010. Graphical methods for investigating the finite-sample properties of confidence regions. Computational Statistics and Data Analysis 54, 262–271.
Phillips, P.C.B., Shimotsu, K., 2004. Local whittle estimation in nonstationary and unit root cases. Annals of Statistics 32, 656–692.
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., Martin, G.M., 2015. Higher-order improvements of the sieve bootstrap for fractionally integrated processes. Journal of Econometrics 188, 94–110.
Robinson, P.M., 1994. Semiparametric analysis of long memory time series. The Annals of Statistics 22, 515–539.
Robinson, P.M., 1995a. Log-periodogram regression of time series with long-range dependence. The Annals of Statistics 23, 1048–1072.
Robinson, P.M., 2005. The distance between rival nonstationary fractional processes. Journal of Econometrics 128, 283–300.
Shao, X., Wu, W.B., 2007. Local whittle estimation of fractional integration for nonlinear processes. Econometric Theory 23, 899–929.
Silva, E.M., Franco, G.C., Reisen, V.A., Cruz, F.R.B., 2006. Local bootstrap approaches for fractional differential parameter estimation in ARFIMA models. Computational Statistics & Data Analysis 51, 1002–1011.
Velasco, C., 1999a. Non-stationary log-periodogram regression. Journal of Econometrics 91, 325–371.
Velasco, C., 1999b. Gaussian semiparametric estimation of non-stationary time series. Journal of Time Series Analysis 20, 87–127.
Velasco, C., 2000. Non gaussian log-periodogram regression. Econometric Theory 16, 44–79.