Frequency Domain Local Bootstrap in long memory time series

Josu Arteche*
Dept. of Quantitative Methods
University of the Basque Country UPV/EHU
Bilbao 48015
Spain
email: josu.arteche@ehu.eus

October 12, 2020

Abstract

Bootstrap techniques in the frequency domain have been proved to be effective instruments to approximate the distribution of many statistics of weakly dependent (short memory) series. However their validity with long memory has not been analysed yet. This paper proposes a Frequency Domain Local Bootstrap (FDLB) based on resampling a locally studentised version of the periodogram in a neighbourhood of the frequency of interest. A bound of the Mallows distance between the distributions of the original and bootstrap periodograms is offered for stationary and non-stationary long memory series. This result is in turn used to justify the use of FDLB for some statistics such as the average periodogram or the Local Whittle (LW) estimator. Finally, the finite sample behaviour of the FDLB in the LW estimator is analysed in a Monte Carlo, comparing its performance with rival alternatives.

Running head: Frequency Domain Local Bootstrap

*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 also thank Jonas Krampe for providing the code of one of the bootstraps used in the simulations.

1 Introduction

Bootstrap is nowadays one of the most popular tools to approximate the distribution of statistics that are unknown or difficult to handle. Originally proposed for iid observations, its application in time series requires dealing with the possible dependence existing between distant observations. Some techniques in the time domain such as the sieve or the block bootstrap have been traditionally used with this purpose. These approaches are based on finally resampling approximately independent quantities: residuals in the sieve or blocks of observations in the block bootstrap. But the strong persistence inherent in long memory time series may invalidate the use of some of these strategies because those quantities may be far from being independent. See for example Lahiri (1993) for the case of the block bootstrap. Different techniques and modifications of some traditional bootstraps have been proposed and their use in long memory series has been theoretically justified in a handful of papers: Andrews et al. (2006) show that a parametric bootstrap based on resampling residuals is effective to construct confidence intervals for maximum likelihood estimators of stationary ARFIMA processes; Poskitt (2008) demonstrates the validity of the sieve bootstrap for linear stationary long memory; and Kim and Nordman (2011) extend Lahiri's results to show the validity of some block bootstrap for the distribution of the sample mean of stationary long memory series. More recently, it has been proposed to apply the bootstrap not on the original observations but rather on the fractionally differenced series $(1-L)^{\hat{d}}x_t$ for $\hat{d}$ a consistent estimator of the memory parameter. This "de-colouring" approach intends to overcome the problems that the strong persistence may originate by getting rid of it with the prior filtering. The final bootstrap series are obtained by "re-colouring" integrating back the resampled series. This strategy has broadened the range of processes and bootstraps that can be used, covering stationary and non-stationary long memory and many of the traditional bootstrap techniques that have been shown to be valid in short memory series (see Poskitt et al. 2015, and Kapetanios et al. 2019). In practice, however, prior fractional differencing and posterior fractional integration are subject to truncation determined by the sample size, which may significantly distort the performance of this approach, especially in small sample sizes.

In addition to the previous strategies in the time domain, some alternative procedures rely on resampling in the frequency domain, which is specially adequate in the context of time series because the Fourier transform converts autocorrelation into heteroscedasticity such that periodogram ordinates of stationary series are asymptotically uncorrelated. Franke and Hardle (1992) and Dahlhaus and Janas (1996) make use of this characteristic to propose a bootstrap strategy for short memory series based on resampling periodogram ordinates studentised with a consistent estimator of the spectral density. The validity of this proposal relies on the consistency of the estimation of the spectral density function, which for short memory time series is simple and well documented but it is especially difficult with long memory series where traditional techniques usually fail (Arteche, 2015). Kim and Nordman (2013) suggest normalising the periodogram with a plug-in estimator of the spectral density when it belongs to a parametric class of functions that covers stationary long memory. They demonstrate that resampling studentised periodograms leads to a consistent estimation of 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 its 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. To the author's knowledge no other bootstrap strategy in the frequency domain has been theoretically justified for long memory series.

This paper focuses on frequencies close to the spectral pole, which are the frequencies where the long memory has its impact and define the region that contains useful information on the persistence of the series. Only the spectral behaviour at those frequencies is restrained, thus avoiding the need for parametric restrictions at frequencies far from the origin. Taking into account the difficulties to estimate non-parametrically the spectral density of long memory series (Arteche, 2015 and Kim and Nordman, 2013), we propose to normalise the periodogram with an estimation of the local behaviour of the spectral density function around the spectral pole to form the locally studentised periodogram (LSP). This approach is similar to the proposal by Franke and Hardle (1992) and Dahlhaus and Janas (1996) but instead of resampling studentised periodograms we propose to resample the LSP. The normalisation in the LSP accounts for the strong persistence but the remaining short memory generates a particular structure in the LSP that should be mimicked by the bootstrap samples, which invalidates the use of a global resampling over the whole band of Fourier frequencies as proposed in Franke and Hardle (1992) and Dahlhaus and Janas (1996). This motivates the use of a local bootstrap scheme based on resampling LSP ordinates in a neighbourhood of the frequency of interest, guaranteeing in that way that the global structure of the LSP is mimicked by the locally bootstrapped LSP.

The rest of the paper is organised as follows. Section 2 describes the characteristics of the long memory processes we deal with. Stationary and non-stationary values of the memory parameter are allowed. Section 3 introduces the Frequency Domain Local Bootstrap (FDLB), shows its validity to approximate first moments of the periodogram at any

Fourier frequency in $(0, \pi]$ and offers a bound of the distance between the true and the FDLB distribution of the periodogram at frequencies close to the spectral pole. This result is in turn used in Section 4 to prove the validity of the FDLB to approximate the distribution of weighted averages of periodogram ordinates close to the spectral pole, as for example an estimator of the spectral distribution function or the score of the local Whittle (LW) estimator. Finally, Section 5 analyses the finite sample behaviour of the FDLB to approximate the distribution of the LW estimator in a Monte Carlo study, comparing its performance with other rival strategies. All the proofs are relegated to Appendix A, whereas Appendix B contains additional Monte Carlo results.

2 Long memory processes

We focus on Long Memory processes x_t satisfying the following assumptions:

A.1: Let d_0 denote the memory parameter of x_t. If $-1/2 < d_0 < 1/2$ then $x_t = v_t$ and for $1/2 \leq d_0 < 1$ then $x_t = x_0 + \sum_{s=1}^{t} v_s$ where x_0 is a random variable not depending on t and $v_t = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j}$, $\sum_{j=0}^{\infty} b_j^2 < \infty$ where $E(\varepsilon_t|F_{t-1}) = 0$, $E(\varepsilon_t^2|F_{t-1}) = 1$, $E(\varepsilon_t^3) < \infty$, $E(\varepsilon_t^4) < \infty$ where F_{t-1} is the σ-field of events generated by ε_s, $s \leq t-1$.

A.2: The spectral density of v_t is

$$f_v(\lambda) = \lambda^{-2d_v} g_v(\lambda) \ , \quad 0 < \lambda \leq \pi \ ,$$

where $-1/2 < d_v < 1/2$ in the stationary case and $-1/2 \leq d_v < 0$ for a nonstationary x_t and $g_v(\lambda)$ is a function which is positive, finite, symmetric around the origin and twice continuously differentiable.

Assumption A.1 avoids the restriction of Gaussianity and only imposes linearity of v_t with bounded fourth moments of the innovations. Nonstationarity is considered as in Velasco (1999) as Type I long memory[1]. The pseudo spectral density function of x_t is in this case $f_x(\lambda) = |1 - \exp(i\lambda)|^{-2} f_v(\lambda)$ such that in both the stationary and non-stationary cases the spectral or pseudo spectral density function of x_t satisfies $f_x(\lambda) = g_v(0)\lambda^{-2d_0}(1 + O(\lambda^2))$ as $\lambda \to 0$ for $d_0 = d_v$ in the stationary case and $d_0 = d_v + 1$ if x_t is nonstationary, entailing $d_0 \in (-1/2, 1)$. Assumption A.2 constrains the possibility of seasonal or cyclical long memory and only allows for standard long memory at frequency zero. The analysis could be extended to cover other types of long memory where the spectral density diverges at a positive frequency as in Arteche and Robinson (2000) but it is constrained here to the empirically more popular case of standard long memory. Note

[1] Kapetanios et al. (2019) consider instead type II long memory series to cover non-stationary cases.

also that the class of models satisfying A.1 and A.2 is wider than previously considered because it includes fractionally integrated processes as particular cases but it also covers more complex structures not relying on fractional differencing, for example those based on cross-sectional aggregation, fractional Gaussian noises (increments of fractional Brownian motions) or self-similar processes.

3 Frequency Domain Local Bootstrap (FDLB)

Many interesting statistics as those discussed in the next section are functions of the periodogram of x_t, which for $t = 1, 2, ..., T$ is defined as

$$I_x(\lambda) = |W_x(\lambda)|^2 \ , \ \ W_x(\lambda) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} x_t \exp(-it\lambda)$$

with $W_x(\lambda)$ being the discrete Fourier transform at frequency λ. The FDLB is here proposed to obtain bootstrap replicates of $I_x(\lambda_j)$ for Fourier frequencies of the form $\lambda_j = 2\pi j/T$, $j = 1, 2, ..., m$, with m satisfying the conditions specified below. Denote $I_j = I_x(\lambda_j)$ and $f_{x,j} = f_x(\lambda_j)$. The FDLB consists of the following steps:

1. Estimate d_0, say $\hat{d}$, and construct the locally studentised periodogram (LSP) $\hat{v}_j = I_j \lambda_j^{2\hat{d}}$, for $j = 1, ..., [T/2]$.

2. Select a resampling width $k_T \in \mathcal{N}$, $k_T \leq [T/2]$.

3. Define i.i.d. discrete random variables $S_1, ..., S_m$ 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).

4. Generate B bootstrap LSP series $\hat{v}_{bj}^* = \hat{v}_{|j+S_j|}$ for $b = 1, 2, ..., B$ and $j = 1, ..., m$.

5. Generate B bootstrap samples for the periodogram $I_{bj}^* = \lambda_j^{-2\hat{d}} \hat{v}_{bj}^*$, for $b = 1, 2, ..., B$ and $j = 1, ..., m$.

6. The bootstrap distribution of the periodogram is calculated as the empirical distribution of the B bootstrap replicates,

$$F_j^*(x) = \frac{1}{B} \sum_{b=1}^{B} \mathbb{I}(I_{bj}^* \leq x)$$

for $\mathbb{I}()$ the indicator function.

Step 1 has some similarities with the de-colouring approach in Poskitt et al. (2015) and Kapetanios et al. (2019) in what its purpose is to reduce the effect of the strong persistence and to get closer to the form of the periodogram of a short memory process. The LSP intends to achieve this goal by dividing by an estimation of the local behaviour of the spectral density function (apart from a constant) in Assumption A.2, whereas Poskitt et al. (2015) and Kapetanios et al. (2019) use fractional differencing. However, the behaviour of $\hat{v}_j$ may still show some structure due to the remaining weak dependence and a global resampling scheme should be avoided. Steps 2 and 3 propose instead a local approach and delimit the band of frequencies where to locally resample. The bootstrap replicates of the periodogram I^*_{bj} are then obtained by resampling $\hat{v}_j$ in a neighbourhood of every frequency λ_j, maintaining in that way the global structure of the periodogram. Note that Δ_T in Step 3 excludes the value $\{-j\}$ to avoid evaluation of the periodogram and of the local approximation of the (pseudo)spectral density at frequency zero, implying that $\#\Delta_T = 2k_T$ if $k_T \geq j$ and $\#\Delta_T = 2k_T + 1$ if $k_T < j$.

Paparoditis and Politis (1999) propose a similar local bootstrap strategy but applied to the raw periodogram I_j instead of to the LSP $\hat{v}_j$. This was shown to be valid under short memory, in a sense that the distance between the original and bootstrap distributions vanishes, but when applied to long memory series no theoretical justification exists. In addition, Silva et al. (2006) show with a Monte Carlo analysis that the resampling width k_T should be very small (as low as $k_T = 1$ or 2) to get sensible results when applied on strongly persistent series, which offers some doubts on the validity of this strategy under long memory. We avoid this problem by locally studentising the periodogram in Step 1.

The validity of the FDLB requires also the following assumptions:

A.3: $1/k_T + k_T/T \to 0$ as $T \to \infty$. Moreover, the sequence $\{p_i; i \in \Delta_T\}$ satisfies

$$\sum_{i\in\Delta_T} p_i = 1, \qquad p_i = p_{-i}, \qquad p_i \to 0, \qquad \sum_{i\in\Delta_T} p_i^2 = O(k_T^{-1}) \text{ as } T \to \infty$$

A.4: As $T \to \infty$

$$\frac{k_T}{T} + \frac{\log|j+k_T|}{k_T^{1/2}} + \frac{\log^{1/2}|j+k_T|}{\overline{jk_T}^{\alpha}} \to 0,$$

where $\alpha = \min\{1/2, 1-d_0\}$ and $\overline{jk_T} = k_T$ if $j/k_T \to 0$ and $\overline{jk_T} = j$ otherwise.

An example of p_i satisfying A.3 is $p_i = 1/\#\Delta_T$. Note that Assumption A.4 implies that $k_T \to \infty$ but j can remain fixed or go to infinity faster or more slowly that k_T, which means that all Fourier frequencies, both close and far from the spectral pole, are covered.

Let E^* denote the expected value calculated with respect to the bootstrap probability conditional on the data $x_1, ..., x_T$. The following theorem offers a bound of the difference between the bootstrap mean of the replicated periodogram ordinates and the spectral density and shows its convergence under the conditions stated in assumption A.4.

Theorem 1 *Under assumptions A.1, A.2 and A.3, if $(\hat{d} - d_0) = O_p(A_T)$ then*

$$E^*(I_j^*) = f_{x,j}\left\{1 + O\left(\left[\frac{k_T}{T}\right]^2\right) + O_p\left(\frac{\log|j+k_T|}{k_T^{1/2}}\mathbb{I}_{(\alpha=1/2)} + \frac{\log^{1/2}|j+k_T|}{jk_T^{\alpha}}\mathbb{I}_{(\alpha<1/2)} + A_T\log T\right)\right\}$$

as $T \to \infty$ uniformly in $j = 1, 2, ..., m$, $m \leq [T/2]$. Additionally, under assumption A,4 and if $A_T = o(\log^{-1} T)$ then $E^*(I_j^*) \xrightarrow{p} f_{x,j}$.

Theorem 1 shows that the FDLB is successful in replicating the mean of the periodogram. In order to analyse further the distributional similarities between I_j^* and I_j we use the Mallows distance between the bootstrap distribution and the true distribution. Consider the set of distribution functions F_2 such that if $F \in F_2$ then $\int_{-\infty}^{\infty} |x|^2 \mathrm{d}F(x) < \infty$. The Mallows distance between two distributions $F, G \in F_2$ is defined as

$$d_2(F, G) = \inf\left\{E|X - Y|^2\right\}^{1/2}$$

where the infimum is taken over all real-valued random variables X and Y with marginal distributions F and G respectively. In what follows we also write $d_2(X, Y)$ for $d_2(F, G)$ when there is no confusion. Note that 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).

Theorem 1 applies for every Fourier frequency in $(0, \pi]$. We focus hereafter on frequencies close to the origin and bound the Mallows distance between I_j^* and I_j for $j = 1, ..., m$, such that $m/T \to 0$ as $T \to \infty$. In particular assumption **A.4** is strengthen as follows:

A.5: As $T \to \infty$

$$\frac{1}{m} + \frac{\log m}{k_T^{1/2}} + \frac{m}{T} + \frac{k_T}{T} \to 0.$$

The restrictions in m imposed by Assumption A.5 entail that only a degenerating band of Fourier frequencies close to zero can be considered. This rules out the possibility of extending the analysis to important class of statistics such as spectral mean estimators (e.g. sample autocovariances) or ratio statistics (e.g. sample autocorrelations), which are functions of the whole band of Fourier frequencies in $(0, \pi)$. However, frequencies close to the spectral pole contain relevant information to analyse the persistence of the series and important statistics are functions of the periodogram at those frequencies, some of them are analysed in Section 4.

Denote by $\mathcal{L}(X)$ the probability distribution function of a variable X and let the locally standardised periodogram be defined as $v_j^0 = I_j \lambda_j^{2d_0}$. Theorem 2 bounds the Mallows distance between the distributions of $\hat{v}_j^*$ and v_j^0.

Theorem 2 *Under assumptions A.1-A.3 and A.5, if* $(\hat{d} - d_0) = o_p(\log^{-1} T)$*, then as* $T \to \infty$*,*

$$d_2[\mathcal{L}(\hat{v}_j^* | x_1, ..., x_T), \mathcal{L}(v_j^0)] = o_p(1) + O_p\left(\frac{\log^{1/2}(k_T + j)}{j^\alpha}\right)$$

for $j = 1, ..., m$*.*

A straightforward corollary of Theorem 2 is the bound for the Mallows distance between the distributions of I_j^* and I_j:

Corollary 1 *Under the assumptions in Theorem 2*

$$d_2[\mathcal{L}(I_j^* | x_1, ..., x_T), \mathcal{L}(I_j)] = o_p(\lambda_j^{-2d_0}) + O_p\left(\lambda_j^{-2d_0} \frac{\log^{1/2}(k_T + j)}{j^\alpha}\right)$$

for $I_j^* = \lambda_j^{-2\hat{d}} \hat{v}_j^*$ *and* $j = 1, ..., m$*.*

4 Weighted averages of periodogram ordinates

Consider now weighted averages of periodogram ordinates of the form

$$\Phi_m = \sum_{j=1}^{m} \psi_j I_j. \tag{1}$$

These statistics are interesting *per se* or as part of more complicated statistics. Consider for example the following two cases:

Average periodogram: Φ_m with $\psi_j = T^{-1} 2\pi$ is a discretely averaged periodogram based on a degenerating band of frequencies around the spectral pole at the origin (Robinson, 1994, and Lobato and Robinson, 1996). Under stationarity, it estimates the spectral distribution function at λ_m. The average peridogram estimator of the memory parameter proposed by Robinson (1994) is defined as a function of a ratio of discrete average periodograms.

Score of the Local Whittle function: The Local Whittle (LW) estimator of the memory parameter d_0 is obtained by minimizing the function $R(d) = \log\left(m^{-1} \sum \lambda_j^{2d} I_j\right) - m^{-1} 2d \sum \log \lambda_j$. Its asymptotic distribution is determined by the weak convergence of the properly normalized score evaluated at d_0: $\sqrt{m} \partial R(d_0)/\partial d \xrightarrow{d} N(0, 4)$. Now $\sqrt{m} \partial R(d_0)/\partial d$

is asymptotically equivalent to Φ_m in (1) with $\psi_j = 2m^{-1/2} g_v^{-1}(0) \lambda_j^{2d_0} v_j$ for $v_j = \log j - m^{-1}\sum_{k=1}^m \log k$ (see Robinson, 1995, for the stationary case and Velasco, 1999, Phillips and Shimotsu, 2004 or Shao and Wu, 2007, for the nonstationary one).

In these cases it is interesting to analyse if the FDLB can be used to approximate the distributional characteristics of Φ_m. Theorem 3 offers a general bound for the Mallows distance between the bootstrap distribution of $\Phi_m^* = \sum_{j=1}^m \psi_j I_j^*$ and the distribution of Φ_m.

Theorem 3 *Under assumptions A.1-A.3, A.5 and $(\hat{d} - d_0) = O_p(A_T)$, as $T \to \infty$.*

$$
\begin{aligned}
d_2^2[\mathcal{L}(\Phi_m^* | x_1, ..., x_T), \mathcal{L}(\Phi_m)] &= O_p\left(\left[\sum |\psi_j| \lambda_j^{-2d_0} B(j, k_T, T)\right]^2\right) \\
&+ \sum \psi_j^2 \lambda_j^{-4d_0}\left[o_p(1) + O_p\left(\frac{\log(j + k_T)}{j^{2\alpha}}\right) + o(1) + O\left(\frac{\log j}{j^{2\alpha}}\right)\right] \\
&+ \sum_j \sum_{l>j} \psi_j \psi_l \lambda_j^{-2d_0} \lambda_l^{-2d_0} O\left(\frac{\log^2 m}{j^{4\alpha}} + \frac{\log^{1/2} m}{j^{2\alpha} l^{\alpha}}\right)
\end{aligned}
$$

where

$$
B(j, k_T, T) = A_T \log T + \frac{\log(j + k_T)}{k_T^{1/2}} \mathbb{I}_{(\alpha=1/2)} + \frac{\log^{1/2}|j + k_T|}{j k_T^{\alpha}} \mathbb{I}_{(\alpha<1/2)} + \frac{k_T^2}{T^2} + \frac{j^2}{T^2}.
$$

When $\psi_j = T^{-1} 2\pi$ and $d_0 < 1/2$ the discrete average periodogram Φ_m is an estimator of the spectral distribution function at λ_m, which can be used for example to estimate the memory parameter in stationary long memory series (Robinson, 1994, and Lobato and Robinson, 1996). The validity of the FDLB in this case is shown in the next corollary, which is a direct application of Theorem 3 under the following additional assumption:

A.6: As $T \to \infty$,

$$
\frac{\log(m + k_T)}{k_T^{\alpha}} + \frac{\log(m + k_T) \log m}{m^{\alpha}} \to 0.
$$

Corollary 2 *Let $\psi_j = T^{-1} 2\pi$. Under assumptions A.1-A.3, A.5, A.6 and $(\hat{d} - d_0) = o_p(\log^{-1} T)$ with $d_0 < 1/2$, as $T \to \infty$,*

$$
d_2^2[\mathcal{L}(\Phi_m^* | x_1, ..., x_T), \mathcal{L}(\Phi_m)] = o_p(1)
$$

.

Note that Corollary 2 justifies the use of the FDLB to approximate the distributional characteristics of Φ_m only when the long memory series is stationary, in which case Φ_m

estimates the spectral distribution function at λ_m. This result can be used also to justify the use of the FDLB to approximate the distribution of the average peridogram estimator of the memory parameter, whose limiting distribution was obtained by Lobato and Robinson (1996) for stationary long memory series.

The validity of the FDLB with the score of the LW function requires the following assumption on k_T and m:

A.7: As $T \to \infty$,

$$\frac{\sqrt{m}\log(m+k_T)\log m}{k_T^\alpha} + \frac{\log(m+k_T)\log^2 m}{m^{2\alpha}} + \frac{\sqrt{m}k_T^2\log m}{T^2} \to 0.$$

Since $\alpha \leq 1/2$ assumption A.7 implies that $\sqrt{m/k_T} \to 0$ and thus $\overline{jk_T} = k_T$ in assumption $A.4$. Note also that the condition $m/k_T \to 0$ in A.7 entails $m^{5/2}T^{-2}\log m \to 0$, which is the condition required in Robinson (1995) and Velasco (1999) for the asymtotic normality of the normalized LW score.

Corollary 3 *Let* $\psi_j = 2m^{-1/2}g_v^{-1}(0)\lambda_j^{2d_0}v_j$ *for* $v_j = \log j - m^{-1}\sum_{k=1}^m \log k$. *Under assumptions A.1-A.3, A.5, A.7 and* $(\hat{d} - d_0) = o_p(m^{-1/2}\log^{-1} T \log^{-1} m)$ *with* $d_0 < 3/4$, *as* $T \to \infty$,

$$d_2^2[\mathcal{L}(\Phi_m^*|x_1,...,x_T), \mathcal{L}(\Phi_m)] = o_p(1).$$

Corollary 3 justifies the validity of the FDLB for the score of the LW function for $d_0 < 3/4$. Note that these are the values for which the score is asymptotically $N(0,4)$ leading to an asymptotically normal distribution of the LW estimator. For $d_0 \geq 3/4$ the score is not asymptotically normal leading to different asymptotic distributions of the LW estimator (see Phillips and Shimotsu, 2004 or Shao and Wu, 2007).

The condition on the rate of convergence of the initial estimator $\hat{d}$ is satisfied by many semiparametric estimators such as the log-periodogram regression or the LW estimator. Since these estimators are $O_p(\sqrt{m_1})$ for a bandwidth m_1, the condition on the corollary entails $m_1^{-1/2}m^{1/2}\log T \log m \to 0$ as $T \to \infty$, which implies that m_1 should be larger than m for large enough T.

5 FDLB in LW estimation: Finite sample performance

One of the main applications of the FDLB is the approximation of the distribution of semiparametric estimators of the memory parameter based on the local behaviour of the spectral density at frequencies close to the spectral pole. Among all the existing techniques, the LW estimator is one of the most popular due to its well known asymptotic properties and excellent finite sample performance. The applicability of the FDLB strategy for the LW estimator was theoretically discussed in the previous section. This section focuses on its validity to approximate the finite sample distribution. Its performance is compared with the following alternatives:

- The asymptotic distribution obtained by Robinson (1995) for stationary processes and Velasco (1999) and Shao and Wu (2007) for type I non-stationary long memory processes as those described in Assumption A.1. When $-1/2 < d_0 < 3/4$ the LW estimator $\hat{d}$ is asymptotically normally distributed as

$$\sqrt{m}(\hat{d} - d_0) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{4}\right)$$

  when the bandwidth m satisfies $m^{-1} + m^5 n^{-4} \log^2 m \to 0$ as $T \to \infty$. Instead of the variance $(4m)^{-1}$ we consider the Hessian-based approximation $\left(4\sum_{j=1}^{m}\left[\log\lambda_j - m^{-1}\sum_{k=1}^{m}\log\lambda_k\right]^2\right)^{-1}$, which has been shown to be significantly closer to the true variance in finite samples (see Hurvich and Chen, 2000 and Arteche, 2006).

- The sieve bootstrap (SBS) proposed by Poskitt (2008), which is based on fitting an AR model of a sufficiently large order. In practice, this order of the AR is selected using AIC and the parameters estimated by Burg's algorithm. Poskitt (2008) proves the validity of this strategy with weak dependent and fractionally integrated processes with $d_0 < 1/2$.

- The prefiletered sieve bootstrap (PFSBS) proposed by Poskitt et al. (2015). This strategy is based on "de-colouring" the original series using the filter $(1 - L)^{\hat{d}}$ where $\hat{d}$ is a consistent estimator of d_0. The SBS is then applied to the filtered series and the final bootstrap samples obtained by fractional integration. Poskitt et al. (2015) prove its validity for fractionally integrated processes with $d_0 < 1/2$, obtaining a faster convergence of the bootstrap distribution than the SBS. Kapetanios et al. (2019) extend this results to non-stationary type II fractional integration. We analyse here its performance in type I long memory.

- The prefiltered spectral-density-driven bootstrap (PFSDDB) based on applying the spectral-density-driven bootstrap proposed by Krampe et al. (2018) to the fractionally differenced series $(1 - L)^{\hat{d}} X_t$. Kapetanios et al. (2019) justify this strategy for type II fractionally integrated series and recommend it over other alternatives such as the block bootstrap based on its finite sample performance. We extend here the analysis to type I long memory.

The comparison is based on the results obtained in 1000 ARFIMA(1,d_0,0) series of the form

$$(1-\phi L)(1-L)^{d_0} X_t = u_t, \quad t = 1, 2, ..T, \tag{2}$$

for $-1/2 < d_0 < 1/2$, where the u_t are standard independent normal. For $d_0 \geq 1/2$ the series is obtained as described in Assumption A.1 by integration of an $ARFIMA(1, d_0-1, 0)$ process. Two values of the autoregressive parameter $\phi = 0$ and $\phi = 0.6$ are considered. FDLB, PFSDDB and PFSBS are based on a preliminary estimation of d_0 obtained by LW with an initial bandwidth m_1.

A simple illustration of the performance of the bootstrap to approximate the finite sample distribution of the LW estimator is offered in Figure 1. It shows kernel estimates of the probability density function of $\hat{d} - d_0$ obtained with 1000 LW estimates of d_0 (true density) together with density estimates of $\hat{d}^b - \hat{d}$ obtained with $b = 1, ..., 1000$ bootstrap replicates in four series with $T = 64$ and $T = 128$. The PFSDDB is not included for the sake of visual presentation and because its performance is similar to the PFSBS, as shown below. The series are generated from the model in equation (2) with $\phi = 0$ and $d_0 = 0.4$, 0.7 and the LW estimates are obtained with $m = 8$ for $T = 64$ and $m = 15$ for $T = 128$. The same values $m_1 = m$ have been used for the initial estimate in the PFSBS and the FDLB. Note that this m_1 does not satisfy the condition in Corollary 3 but, according to the Monte Carlo below, this is a valid option. The resampling width used in the FDLB is $k_T = 10$ for $T = 64$ and $k_T = 20$ for $T = 128$.

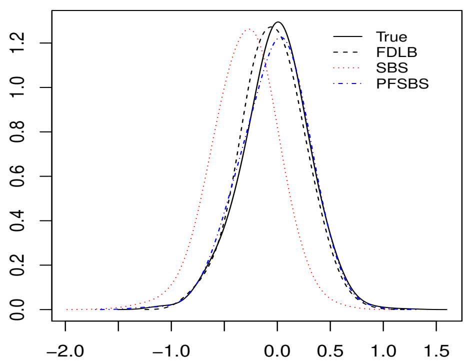
Figure 1 shows that the three bootstrap strategies can be effective to approximate the distribution of the LW estimator, the FDLB offering better approximations than the PFSBS and the SBS, at least for the four series considered. In order to give a more general picture not only based on four individual realizations, two different comparative global measures are considered: the root mean squared deviation of the bootstrap probability density function to the true (obtained with Monte Carlo) density, and the coverage frequencies of confidence intervals obtained with the asymptotic distribution and the bootstrap estimates in different models. Both measures are obtained with $R = 1000$ simulations of fractional Gaussian noises with memory parameters $d_0 = -04$, 0.4 and 0.7.

The root mean squared deviation is defined as

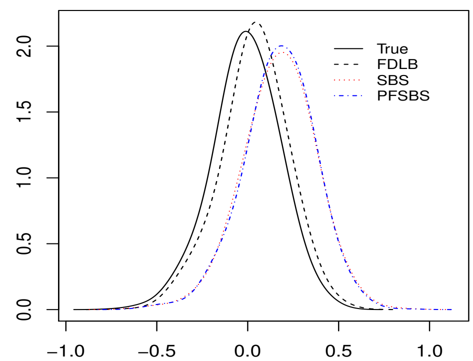
$$RMSD(boot) = \sqrt{\frac{1}{R}\sum_{j=1}^{R}\left(p_{boot}(\hat{d}_j - d_0) - p_{mc}(\hat{d}_j - d_0)\right)^2}$$

Figure 1: True and bootstrap distributions

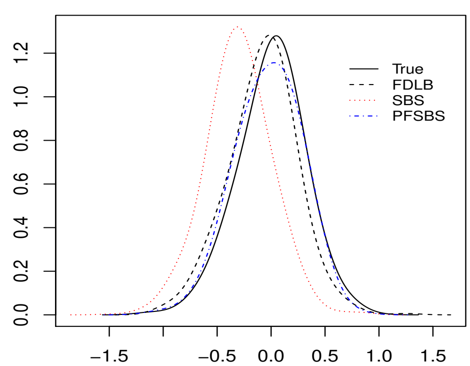
(a) $d_0 = 0.4$, $T = 64$, $m = 8$, $m_1 = 8$, $k_T = 10$



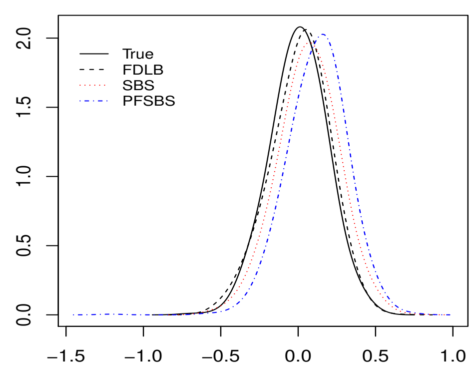
(b) $d_0 = 0.4$, $T = 128$, $m = 15$, $m_1 = 15$, $k_T = 20$



(c) $d_0 = 0.7$, $T = 64$, $m = 8$, $m_1 = 8$, $k_T = 10$



(d) $d_0 = 0.7$, $T = 128$, $m = 15$, $m_1 = 15$, $k_T = 20$



Note: Monte Carlo (1000 replications) and bootstrap (1000 bootstrap samples) distributions for $(1-L)^{d_0}X_t = u_t$, $t = 1, ...T$, $u_t \sim \mathcal{N}(0,1)$.

where *boot* represents one of the bootstrap strategies discussed above, R is the number of simulations, $p_{mc}(\hat{d}_j - d_0)$ is the ordinate of a kernel based density estimate obtained with the R estimates $\hat{d}_k - d_0$, $k = 1, ..., R$ 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 $\hat{d}_j^b - \hat{d}_j$, $b = 1, ..., B$ evaluated at $\hat{d}_j - d_0$. The numbers in Table 1 are ratios of the RMSD(PFSBS), RMSD(PFSDDB) and RMSD(FDLB) over RMSD(SBS) for $\phi = 0$, hence a value less than one indicates a better fit of the bootstrap in the numerator than the SBS. In general the PFSBS, the PFSDDB and the FDLB offer better fits than the SBS, and only in very few cases the ratio is larger than one, mainly when the bandwidth in the initial estimate of d_0 used to pre-filter (m_1) is lower than the bandwidth employed in the final estimate (m). A value $m_1 \geq m$ is generally a wiser option, which is in agreement with the condition in Corollary 3. In 9 out of the 18 cases the FDLB with an appropriate combination of m_1 and k_T gives the lowest ratio, whereas in the other 9 the PFSDDB is the best option. As a rule m_1 should be close to but larger than m. Regarding the selection of k_T, no definitive suggestion can be extracted from Table 1. The only conclusion that seems clear is that a large k_T is significantly more harmful than a lower one when $m_1 < m$. However, when $m_1 \geq m$ a larger k_T leads to better results.

Tables 2, 3 and 4 show coverage frequencies and average widths of confidence intervals obtained with $R = 1000$ replications of fractional Gaussian noises with memory parameters $d_0 = -04$, 0.4 and 0.7 respectively, which are values covered by the theoretical results in the previous section. Note however that the theoretical validity of the SBS and PFSB was not originally discussed for nonstationary long memory series as those obtained with $d = 0.7$. It was only recently when Kapetanios et al. (2019) prove that prefiltering is a valid option also for nonstationary long memory as long as a consistent estimator of d exists, but their results only apply to type II long memory processes instead of type I long memory as the processes considered here.

Tables 2-4 show that the asymptotic distribution tends to lead to under-coverage, although the use of the Hessian-based approximation significantly enlarges the confidence intervals and improves the coverage (see Arteche and Orbe, 2016). The different bootstrap strategies tend to improve the coverages, with the FDLB leading in 13 out of 18 times to the closest-to-nominal coverage for some of the values of m_1 and k_T, even with narrower intervals. In general, increasing k_T leads to higher coverages with wider intervals, with some over-coverages. As a rule a larger k_T leads to better results for large values of d_0, which is in agreement with the condition in assumption A.7 for the evolution of k_T. The selection of m_1 is also important. Those cases with $m_1 < m$ tends to lead to under-coverage. This is in agreement with the condition imposed in the rate of convergence of the estimator of d_0 used to locally studentise the periodogram in Corollary 3, according to which it should converge faster than $m^{-1/2}$. It also agrees with the results concerning the RMSD in Table 1, which shows that $m_1 \geq m$ is a wise choice not only for the FDLB but also for the the PFSBS and PFSDDB.

All the results shown so far come from fractional noises with no additional weak dependence. The conclusions with an ARFIMA(1,d,0) with $\phi = 0.6$ are somewhat similar, as shown in the results in Appendix B. The FDLB offers now the lowest RMSD in 10 out of 18 cases and the closest coverage to the nominal 95% in 13 out of 18. The main difference with respect to the results with fractional noises is the twofold biasing effect caused by

Table 1: Ratio of RMSD: ARFIMA(0,d_0,0)

	$d_0 = -0.4$, $T = 64$								
	$m = 3$			$m = 8$			$m = 15$		
	$m_1 = 3$	$m_1 = 8$	$m_1 = 15$	$m_1 = 3$	$m_1 = 8$	$m_1 = 15$	$m_1 = 3$	$m_1 = 8$	$m_1 = 15$
PFSBS	0.593	0.684	0.707	0.888	0.369	0.713	0.899	0.653	0.403
PFSDDB	0.608	0.722	0.761	1.191	0.270	0.729	1.524	1.025	**0.300**
FDLB ($k_T = 3$)	1.748	**0.577**	0.702	0.417	0.589	0.309	0.372	0.653	0.649
FDLB ($k_T = 5$)	1.431	0.615	0.641	0.778	0.447	0.307	0.599	0.469	0.468
FDLB ($k_T = 10$)	0.996	0.591	0.641	1.476	0.330	0.640	1.738	0.547	0.375
FDLB ($k_T = 25$)	0.783	0.640	0.680	1.681	**0.266**	0.636	2.410	1.199	0.325
	$d_0 = -0.4$, $T = 128$								
	$m = 5$			$m = 15$			$m = 30$		
	$m_1 = 5$	$m_1 = 15$	$m_1 = 30$	$m_1 = 5$	$m_1 = 15$	$m_1 = 30$	$m_1 = 5$	$m_1 = 15$	$m_1 = 30$
PFSBS	0.276	0.629	0.677	0.609	0.219	0.483	0.738	0.622	0.456
PFSDDB	**0.230**	0.629	0.676	0.882	**0.171**	0.489	1.440	0.867	**0.338**
FDLB ($k_T = 5$)	0.515	0.386	0.427	0.332	0.354	0.252	0.646	0.838	0.797
FDLB ($k_T = 10$)	0.351	0.437	0.514	0.709	0.291	0.353	0.712	0.736	0.637
FDLB ($k_T = 20$)	0.268	0.519	0.578	1.146	0.276	0.482	1.737	0.665	0.572
FDLB ($k_T = 50$)	0.324	0.612	0.640	1.246	0.255	0.456	2.332	1.041	0.563
	$d_0 = 0.4$, $T = 64$								
	$m = 3$			$m = 8$			$m = 15$		
	$m_1 = 3$	$m_1 = 8$	$m_1 = 15$	$m_1 = 3$	$m_1 = 8$	$m_1 = 15$	$m_1 = 3$	$m_1 = 8$	$m_1 = 15$
PFSBS	0.501	0.962	0.975	0.951	0.349	0.705	1.440	0.838	0.386
PFSDDB	0.350	1.003	1.079	1.249	0.275	0.737	2.208	1.403	**0.256**
FDLB ($k_T = 3$)	2.224	0.634	0.578	0.317	0.535	**0.251**	0.382	0.695	0.735
FDLB ($k_T = 5$)	1.581	0.560	0.645	0.792	0.403	0.299	0.707	0.356	0.501
FDLB ($k_T = 10$)	0.656	0.633	0.764	1.483	0.277	0.646	2.191	0.577	0.326
FDLB ($k_T = 25$)	**0.349**	0.898	0.934	1.697	0.255	0.666	3.050	1.467	0.263
	$d_0 = 0.4$, $T = 128$								
	$m = 5$			$m = 15$			$m = 30$		
	$m_1 = 5$	$m_1 = 15$	$m_1 = 30$	$m_1 = 5$	$m_1 = 15$	$m_1 = 30$	$m_1 = 5$	$m_1 = 15$	$m_1 = 30$
PFSBS	0.293	0.790	0.832	0.797	0.178	0.542	1.314	0.847	0.399
PFSDDB	0.231	0.793	0.842	1.154	**0.199**	0.555	2.251	1.456	0.379
FDLB ($k_T = 5$)	0.475	0.498	0.548	0.341	0.450	0.311	0.329	0.556	0.559
FDLB ($k_T = 10$)	**0.189**	0.569	0.659	0.835	0.330	0.373	0.847	0.335	**0.316**
FDLB ($k_T = 20$)	0.298	0.678	0.735	1.407	0.312	0.562	2.520	0.726	0.359
FDLB ($k_T = 50$)	0.446	0.784	0.803	1.507	0.270	0.546	3.331	1.561	0.468
	$d_0 = 0.7$, $T = 64$								
	$m = 3$			$m = 8$			$m = 15$		
	$m_1 = 3$	$m_1 = 8$	$m_1 = 15$	$m_1 = 3$	$m_1 = 8$	$m_1 = 15$	$m_1 = 3$	$m_1 = 8$	$m_1 = 15$
PFSBS	0.589	0.730	0.793	0.763	0.463	0.770	1.369	0.915	0.573
PFSDDB	0.563	0.731	0.789	1.028	0.343	0.737	1.983	1.326	**0.374**
FDLB ($k_T = 3$)	1.195	**0.406**	0.442	0.369	0.453	0.315	0.399	0.636	0.637
FDLB ($k_T = 5$)	0.843	0.493	0.568	0.812	0.344	0.441	0.740	0.421	0.405
FDLB ($k_T = 10$)	0.456	0.585	0.648	1.430	**0.285**	0.705	2.088	0.706	0.396
FDLB ($k_T = 25$)	0.430	0.749	0.752	1.624	0.380	0.732	2.873	1.477	0.486
	$d_0 = 0.7$, $T = 128$								
	$m = 5$			$m = 15$			$m = 30$		
	$m_1 = 5$	$m_1 = 15$	$m_1 = 30$	$m_1 = 5$	$m_1 = 15$	$m_1 = 30$	$m_1 = 5$	$m_1 = 15$	$m_1 = 30$
PFSBS	0.292	0.704	0.737	1.000	0.204	0.684	1.952	1.210	0.592
PFSDDB	0.250	0.706	0.743	1.343	**0.141**	0.651	3.115	1.812	**0.439**
FDLB ($k_T = 5$)	0.327	0.499	0.516	0.381	0.519	0.349	0.657	1.028	1.036
FDLB ($k_T = 10$)	**0.155**	0.538	0.582	1.090	0.399	0.427	1.122	0.791	0.742
FDLB ($k_T = 20$)	0.276	0.641	0.671	1.806	0.394	0.698	3.282	1.054	0.894
FDLB ($k_T = 50$)	0.419	0.742	0.744	1.931	0.379	0.713	4.420	2.040	1.057

Note: The numbers in each cell show the ratio of RMSD obtained with the different bootstrap strategies with respect to the sieve. In bold the lowest ratio for every m.

the short memory component for large bandwidths. First, a large m induces a bias in the estimation of d_0 that deteriorates the coverage of all the confidence intervals, using both the asymptotic distribution and the bootstrap approximations. Second, a large m_1 worsens the initial estimation of d_0 used in the de-colouring step, which in turn causes a deterioration in the PFSDDB, PFSBS and FDLB approximations of the LW distribution. This implies that $m_1 \geq m$ is not always beneficial because the bias in the initial estimate of d_0 may significantly deteriorate the performance of the filtered approaches. In any case the bootstrap significantly improves the coverage over the asymptotic distribution, in a much greater degree than in the case of fractional noise. This improvement is more evident with the FDLB, where the coverage can go from less than 10% using the asymptotic distribution to more than 60% using the FDLB (see Tables 6-8 in Appendix B).

6 Conclusion

Pre-filtering either in the time domain, as in the PFSBS or the PFSDDB, or in the frequency domain, as in the FDLB, is shown to be a beneficial strategy prior to resampling. The SBS has instead the advantage of being fully automatic because the order of the autoregression can be selected by widely accepted strategies, for example minimising some information criteria as the AIC. The pre-filtered approaches require the intervention of the user to select m_1 and the FDLB needs in addition a prior selection of k_T. However, a combination of m_1 and k_T can be found for which the FDLB performs better than the SBS and in many cases also better than the PFSBS and the PFSDDB. According to the results obtained in the Monte Carlo analysis, selecting m_1 close to (but larger than) m seems a sensible option, at least when there is not a significant short memory component. As far as the selection of k_T is concerned, the Monte Carlo results show that a too large k_T leads to over-coverage when m is small but a too small k_T leads to under-coverage when m is large. A sensible option, at least in terms of coverage, is then to select a value of k_T slightly larger than m but low when m is small and large for bigger m. However a more rigorous analysis is necessary before offering general advices on the choice of both m_1 and k_T and it is left for future research.

References

[1] Andrews, D.W.K., O. Lieberman, and V. Marmer (2006) Higher-order improvements of the parametric bootstrap for long-memory time series. *Journal of Econometrics* 133, 673-702.

Table 2: Coverages and widths of confidence intervals: ARFIMA(0,-0.4,0)

	$T=64$								
	$m=3$			$m=8$			$m=15$		
	$m_1=3$	$m_1=8$	$m_1=15$	$m_1=3$	$m_1=8$	$m_1=15$	$m_1=3$	$m_1=8$	$m_1=15$
Asymp.	**0.959**	**0.959**	**0.959**	0.916	0.916	0.916	0.928	0.928	0.928
	(2.495)	(2.495)	(2.495)	(1.053)	(1.053)	(1.053)	(0.670)	(0.670)	(0.670)
SBS	1.000	1.000	1.000	0.996	0.996	0.996	0.983	0.983	0.983
	(2.663)	(2.663)	(2.663)	(1.228)	(1.228)	(1.228)	(0.770)	(0.770)	(0.770)
PFSBS	0.982	1.000	1.000	**0.942**	0.987	0.987	0.943	0.970	0.979
	(2.396)	(2.398)	(2.394)	(1.216)	(1.224)	(1.227)	(0.769)	(0.760)	(0.758)
PFSDDB	0.983	1.000	1.000	0.934	0.994	0.995	0.851	0.934	0.980
	(2.411)	(2.423)	(2.426)	(1.233)	(1.216)	(1.218)	(0.781)	(0.748)	(0.742)
FDLB($k_T=3$)	**0.941**	0.974	0.995	0.874	0.887	0.939	0.871	0.872	0.873
	(1.890)	(1.942)	(2.027)	(1.028)	(0.969)	(0.987)	(0.671)	(0.614)	(0.624)
FDLB($k_T=5$)	0.983	0.986	0.998	0.911	0.921	0.971	0.898	0.896	0.903
	(2.125)	(2.041)	(2.144)	(1.186)	(1.027)	(1.048)	(0.772)	(0.643)	(0.652)
FDLB($k_T=10$)	0.999	1.000	1.000	0.914	0.964	0.987	0.830	0.916	**0.947**
	(2.352)	(2.231)	(2.251)	(1.303)	(1.115)	(1.091)	(0.894)	(0.701)	(0.675)
FDLB($k_T=25$)	1.000	1.000	1.000	0.971	0.984	0.998	0.627	0.853	0.955
	(2.551)	(2.403)	(2.356)	(1.412)	(1.224)	(1.182)	(0.919)	(0.764)	(0.717)
	$T=128$								
	$m=5$			$m=15$			$m=30$		
	$m_1=5$	$m_1=15$	$m_1=30$	$m_1=5$	$m_1=15$	$m_1=30$	$m_1=5$	$m_1=15$	$m_1=30$
Asymp.	0.908	0.908	0.908	0.909	0.909	0.909	0.915	0.915	0.915
	(1.542)	(1.542)	(1.542)	(0.670)	(0.670)	(0.670)	(0.428)	(0.428)	(0.428)
SBS	1.000	1.000	1.000	0.990	0.990	0.990	0.835	0.835	0.835
	(1.863)	(1.863)	(1.863)	(0.746)	(0.746)	(0.746)	(0.473)	(0.473)	(0.473)
PFSBS	0.992	1.000	1.000	**0.951**	0.982	0.990	0.934	0.947	0.962
	(1.763)	(1.800)	(1.820)	(0.751)	(0.743)	(0.743)	(0.472)	(0.463)	(0.460)
PFSDDB	0.990	1.000	1.000	0.885	0.978	0.997	0.756	0.892	0.970
	(1.755)	(1.804)	(1.822)	(0.755)	(0.743)	(0.740)	(0.472)	(0.460)	(0.457)
FDLB($k_T=5$)	**0.957**	0.988	0.999	0.887	0.918	0.953	0.893	0.897	0.911
	(1.467)	(1.498)	(1.538)	(0.677)	(0.658)	(0.666)	(0.430)	(0.405)	(0.410)
FDLB($k_T=10$)	0.983	0.998	1.000	0.868	0.947	0.983	0.881	0.910	0.929
	(1.637)	(1.587)	(1.634)	(0.766)	(0.679)	(0.686)	(0.483)	(0.420)	(0.424)
FDLB($k_T=20$)	1.000	1.000	1.000	0.763	0.948	0.992	0.692	0.910	0.955
	(1.790)	(1.688)	(1.706)	(0.810)	(0.710)	(0.698)	(0.540)	(0.444)	(0.434)
FDLB($k_T=50$)	1.000	1.000	1.000	0.824	0.960	0.996	0.477	0.821	**0.950**
	(1.909)	(1.780)	(1.772)	(0.869)	(0.755)	(0.731)	(0.559)	(0.469)	(0.450)

Note: The first number in each cell is the coverage frequency obtained with 1000 replications. The number between round brackets is the average width of the confidence intervals. In bold the closest coverage to 0.95 nominal coverage for every m.

[2] Arteche, J. (2006) Semiparametric estimation in perturbed long memory series. *Computational Statistics and Data Analysis* 51, 2118-2141.

[3] Arteche, J. (2015) Signal extraction in Long Memory Stochastic Volatility. *Economeric Theory* 31, 1382-1402.

Table 3: Coverages and widths of confidence intervals: ARFIMA(0,0.4,0)

	$T=64$								
	$m=3$			$m=8$			$m=15$		
	$m_1=3$	$m_1=8$	$m_1=15$	$m_1=3$	$m_1=8$	$m_1=15$	$m_1=3$	$m_1=8$	$m_1=15$
Asymp.	0.880	0.880	0.880	0.914	0.914	0.914	0.912	0.912	0.912
	(2.495)	(2.495)	(2.495)	(1.053)	(1.053)	(1.053)	(0.670)	(0.670)	(0.670)
SBS	0.994	0.994	0.994	0.973	0.973	0.973	0.921	0.921	0.921
	(2.851)	(2.851)	(2.851)	(1.241)	(1.241)	(1.241)	(0.766)	(0.766)	(0.766)
PFSBS	0.984	0.997	0.996	0.902	0.970	0.969	0.885	0.934	**0.946**
	(2.678)	(2.937)	(3.010)	(1.164)	(1.228)	(1.230)	(0.750)	(0.760)	(0.755)
PFSDDB	0.985	1.000	1.000	0.835	**0.957**	0.988	0.715	0.848	0.924
	(2.726)	(2.973)	(3.047)	(1.203)	(1.216)	(1.215)	(0.759)	(0.742)	(0.739)
FDLB ($k_T=3$)	0.921	0.975	0.995	0.851	0.859	0.914	0.867	0.852	0.859
	(2.046)	(2.093)	(2.195)	(1.106)	(1.001)	(1.016)	(0.704)	(0.616)	(0.628)
FDLB ($k_T=5$)	**0.973**	0.986	0.999	0.887	0.903	0.960	0.889	0.864	0.876
	(2.328)	(2.262)	(2.389)	(1.297)	(1.053)	(1.067)	(0.822)	(0.646)	(0.653)
FDLB ($k_T=10$)	1.000	0.999	1.000	0.892	0.938	0.980	0.830	0.875	0.917
	(2.664)	(2.576)	(2.611)	(1.434)	(1.139)	(1.103)	(1.003)	(0.706)	(0.676)
FDLB ($k_T=25$)	1.000	1.000	1.000	0.962	0.967	0.991	0.631	0.836	0.919
	(2.945)	(2.872)	(2.857)	(1.600)	(1.278)	(1.194)	(1.074)	(0.783)	(0.722)
	$T=128$								
	$m=5$			$m=15$			$m=30$		
	$m_1=5$	$m_1=15$	$m_1=30$	$m_1=5$	$m_1=15$	$m_1=30$	$m_1=5$	$m_1=15$	$m_1=30$
Asymp.	0.915	0.915	0.915	0.924	0.924	0.924	0.916	0.916	0.916
	(1.542)	(1.542)	(1.542)	(0.670)	(0.670)	(0.670)	(0.428)	(0.428)	(0.428)
SBS	0.997	0.997	0.997	0.922	0.922	0.922	0.900	0.900	0.900
	(1.884)	(1.884)	(1.884)	(0.750)	(0.750)	(0.750)	(0.471)	(0.471)	(0.471)
PFSBS	0.989	0.999	0.998	0.909	0.964	0.980	0.895	0.924	**0.940**
	(1.836)	(1.874)	(1.878)	(0.745)	(0.739)	(0.739)	(0.478)	(0.463)	(0.459)
PFSDDB	0.989	1.000	1.000	0.793	0.946	0.984	0.700	0.848	0.930
	(1.846)	(1.875)	(1.880)	(0.742)	(0.738)	(0.737)	(0.471)	(0.459)	(0.456)
FDLB($k_T=5$)	**0.957**	0.990	0.997	0.896	0.915	0.946	0.887	0.892	0.896
	(1.589)	(1.603)	(1.627)	(0.679)	(0.659)	(0.662)	(0.431)	(0.406)	(0.409)
FDLB($k_T=10$)	0.983	0.996	1.000	0.873	**0.950**	0.974	0.900	0.916	0.913
	(1.812)	(1.713)	(1.749)	(0.787)	(0.680)	(0.684)	(0.487)	(0.420)	(0.423)
FDLB($k_T=20$)	1.000	1.000	1.000	0.764	0.955	0.985	0.709	0.906	**0.940**
	(1.989)	(1.822)	(1.817)	(0.861)	(0.712)	(0.700)	(0.554)	(0.444)	(0.435)
FDLB($k_T=50$)	1.000	1.000	1.000	0.822	0.958	0.989	0.793	0.820	0.927
	(2.113)	(1.914)	(1.870)	(0.940)	(0.760)	(0.733)	(0.601)	(0.472)	(0.453)

Note: The first number in each cell is the coverage frequency obtained with 1000 replications. The number between round brackets is the average width of the confidence intervals. In bold the closest coverage to 0.95 nominal coverage for every m.

[4] Arteche, J. and J. Orbe (2016) A bootstrap approximation for the distribution of the Local Whittle estimator. *Computational Statistics and Data Analysis 100*, 645-660

[5] Arteche, J. and P.M. Robinson (2000) Semiparametric inference in seasonal and cyclical long memory processes. *Journal of Time Series Analysis* 21, 1-27.

[6] Bickel, P. J. and D. A. Freedman (1981) Some asymptotic theory for the bootstrap.

Table 4: Coverages and widths of confidence intervals: ARFIMA(0,0.7,0)

	$T=64$								
	$m=3$			$m=8$			$m=15$		
	$m_1=3$	$m_1=8$	$m_1=15$	$m_1=3$	$m_1=8$	$m_1=15$	$m_1=3$	$m_1=8$	$m_1=15$
Asymp.	0.940	0.940	0.940	0.916	0.916	0.916	0.942	0.942	0.942
	(2.495)	(2.495)	(2.495)	(1.053)	(1.053)	(1.053)	(0.670)	(0.670)	(0.670)
SBS	0.987	0.987	0.987	0.956	0.956	0.956	**0.945**	**0.945**	**0.945**
	(2.951)	(2.951)	(2.951)	(1.233)	(1.233)	(1.233)	(0.759)	(0.759)	(0.759)
PFSBS	0.995	0.993	0.997	0.935	0.964	0.981	0.893	0.932	0.955
	(2.602)	(2.833)	(2.914)	(1.139)	(1.199)	(1.209)	(0.726)	(0.744)	(0.746)
PFSDDB	0.997	0.999	0.999	0.854	0.957	0.992	0.731	0.859	0.961
	(2.624)	(2.845)	(2.916)	(1.145)	(1.179)	(1.197)	(0.723)	(0.724)	(0.729)
FDLB ($k_T=3$)	**0.949**	0.984	0.998	0.894	0.909	**0.945**	0.904	0.883	0.884
	(2.013)	(2.033)	(2.117)	(1.092)	(1.008)	(1.020)	(0.687)	(0.614)	(0.625)
FDLB ($k_T=5$)	0.994	0.994	1.000	0.927	0.940	0.975	0.918	0.899	0.915
	(2.309)	(2.216)	(2.327)	(1.275)	(1.067)	(1.079)	(0.805)	(0.647)	(0.655)
FDLB ($k_T=10$)	1.000	1.000	1.000	0.914	0.966	0.985	0.863	0.913	0.940
	(2.628)	(2.533)	(2.560)	(1.392)	(1.145)	(1.109)	(0.968)	(0.708)	(0.678)
FDLB ($k_T=25$)	1.000	1.000	1.000	0.980	0.979	0.994	0.656	0.863	**0.945**
	(2.911)	(2.815)	(2.800)	(1.534)	(1.261)	(1.195)	(1.026)	(0.781)	(0.723)
	$T=128$								
	$m=5$			$m=15$			$m=30$		
	$m_1=5$	$m_1=15$	$m_1=30$	$m_1=5$	$m_1=15$	$m_1=30$	$m_1=5$	$m_1=15$	$m_1=30$
Asymp.	0.921	0.921	0.921	0.925	0.925	0.925	0.921	0.921	0.921
	(1.542)	(1.542)	(1.542)	(0.670)	(0.670)	(0.670)	(0.428)	(0.428)	(0.428)
SBS	0.995	0.995	0.995	0.924	0.924	0.924	**0.948**	**0.948**	**0.948**
	(1.879)	(1.879)	(1.879)	(0.753)	(0.753)	(0.753)	(0.472)	(0.472)	(0.472)
PFSBS	0.994	0.998	0.997	0.905	0.966	0.988	0.889	0.921	0.947
	(1.780)	(1.838)	(1.848)	(0.727)	(0.731)	(0.736)	(0.468)	(0.465)	(0.463)
PFSDDB	0.985	0.999	1.000	0.819	**0.953**	0.996	0.684	0.835	0.936
	(1.779)	(1.836)	(1.848)	(0.720)	(0.727)	(0.733)	(0.455)	(0.455)	(0.458)
FDLB($k_T=5$)	**0.961**	0.992	1.000	0.899	0.900	0.935	0.894	0.881	0.887
	(1.540)	(1.563)	(1.596)	(0.669)	(0.649)	(0.653)	(0.426)	(0.401)	(0.403)
FDLB($k_T=10$)	0.984	0.999	1.000	0.885	**0.947**	0.982	0.892	0.900	0.903
	(1.759)	(1.669)	(1.714)	(0.773)	(0.668)	(0.673)	(0.481)	(0.413)	(0.418)
FDLB($k_T=20$)	1.000	1.000	1.000	0.781	0.956	0.989	0.718	0.903	0.946
	(1.921)	(1.779)	(1.781)	(0.847)	(0.700)	(0.688)	(0.545)	(0.437)	(0.429)
FDLB($k_T=50$)	1.000	1.000	1.000	0.834	0.961	0.993	0.508	0.818	0.931
	(2.063)	(1.888)	(1.861)	(0.930)	(0.759)	(0.731)	(0.600)	(0.469)	(0.453)

Note: The first number in each cell is the coverage frequency obtained with 1000 replications. The number between round brackets is the average width of the confidence intervals. In bold the closest coverage to 0.95 nominal coverage for every m.

Annals of Statistics 9, 1196-1217.

[7] Brockwell, P.J. and R.A. Davis (2006) *Time Series: Theory and Methods*. Springer.

[8] Dahlhaus, R. and D. Janas (1996) A frequency domain bootstrap for ratio statistics in time series analysis. *Annals of Statistics* 24, 1934-1963.

[9] Franke, J. and W. Härdle (1992) On bootstrapping kernel spectral estimates. *Annals of Statistics* 20, 121-145.

[10] Hidalgo, J. (2003) An alternative bootstrap to moving blocks for time series regression models. *Journal of Econometrics* 117, 369-399.

[11] Hurvich, C.M. and W.W. Chen (2000) An efficient taper for potentially overdifferenced long-memory time series. *Journal of Time Series Analysis* 21, 155-180.

[12] Kapetanios, G., F. Papailias and A.M.R. Taylor (2019) A Generalised Fractional Differencing Bootstrap for Long Memory Processes. *Journal of Time Series Analysis* 40, 467-492.

[13] Kim, Y.M. and D.J. Nordman (2011) Properties of a block bootstrap under long-range dependence. *Sankhya Ser A* 73, 79–109.

[14] Kim, Y.M. and D.J. Nordman (2013) A frequency domain bootstrap for Whittle estimation under long-range dependence. *Journal of Multivariate Analysis* 115, 405-420.

[15] Krampe, J., J. P. Kreiss and E. Paparoditis (2018) Estimated Wold representation and spectral-density-driven bootstrap for time series. *Journal of the Royal Statistical Society Series B* 80, 703-726.

[16] Lahiri, S.N. (1993) On the moving block bootstrap under long range dependence. *Statistics and Probability Letters* 18, 405-413.

[17] Lobato, I. and P.M. Robinson (1996) Averaged periodogram estimation of long memory. *Journal of Econometrics* 73, 303-324.

[18] Paparoditis, E. and D. Politis (1999) The local bootstrap for periodogram statistics. *Journal of Time Series Analysis* 20, 193-222.

[19] Phillips, P.C.B. and K. Shimotsu (2004) Local Whittle estimation in nonstationary and unit root cases. *Annals of Statistics* 32, 656-692.

[20] Poskitt, D.S. (2008) Properties of the sieve bootstrap for fractionally integrated and non-invertible processes. *Journal of Time Series Analysis* 29, 224-250.

[21] Poskitt, D.S., S.D. Grose and G.M. Martin (2015) Higher-order improvements of the sieve bootstrap for fractionally integrated processes. *Journal of Econometrics* 188, 94-110.

[22] Robinson, P.M. (1994) Semiparametric analysis of long-memory time series. *Annals of Statistics* 22, 515-539.

[23] Robinson, P.M. (1995) Gaussian semiparametric estimation of long-range dependence. *Annals of Statistics* 23, 1630-1661.

[24] Shao, X. and W.B. Wu (2007) Local Whittle estimation of fractional integration for nonlinear processes. *Econometric Theory* 23, 899-929.

[25] Silva, E.M., G.C. Franco, V.A. Reisen and F.R.B. Cruz (2006) Local bootstrap approaches for fractional differential parameter estimation in ARFIMA models. *Computational Statistics and Data Analysis* 51, 1002-1011.

[26] Velasco, C. (1999) Gaussian semiparametric estimation of non-stationary time series. *Journal of Time Series Analysis* 20, 87-127.

A Appendix: Proofs of the theorems

Proof of Theorem 1:

$$E^*(I_j^*) = \sum_{i\in\Delta_T} p_i \lambda_j^{-2\hat{d}} \hat{v}_{j+i} = \sum_{i\in\Delta_T} p_i \left|1+\frac{i}{j}\right|^{2\hat{d}} I_{j+i} \tag{A.1}$$

and

$$\sum_{i\in\Delta_T} p_i \left|1+\frac{i}{j}\right|^{2\hat{d}} I_{j+i} - f_{x,j} = \lambda_j^{-2d_0} \sum_{i\in\Delta_T} p_i |\lambda_j+\lambda_i|^{2d_0} I_{j+i} - f_{x,j} \tag{A.2}$$

$$+ \sum_{i\in\Delta_T} p_i I_{j+i} \left|1+\frac{i}{j}\right|^{2d_0} \left\{\left|1+\frac{i}{j}\right|^{2(\hat{d}-d_0)} - 1\right\} \tag{A.3}$$

Now the right hand side of (A.2) is equal to

$$\lambda_j^{-2d_0} \sum_{i\in\Delta_T} p_i |\lambda_j+\lambda_i|^{2d_0} (I_{j+i} - f_{x,j+i}) + \lambda_j^{-2d_0} \sum_{i\in\Delta_T} p_i |\lambda_j+\lambda_i|^{2d_0} f_{x,j+i} - f_{x,j} \tag{A.4}$$

By assumption A.2 the second summand in (A.4) is bounded by

$$\lambda_j^{-2d_0} \sum_{i\in\Delta_T} p_i \frac{\lambda_i^2}{2} \max_{\lambda_i\le\lambda\le\lambda_j+\lambda_i} |g_x''(\lambda)| = O\left(\lambda_j^{-2d_0} (\sum p_i^2)^{1/2} (\sum \lambda_i^4)^{1/2}\right)$$

$$= O\left(\lambda_j^{-2d_0} k_T^{-1/2} \frac{k_T^{5/2}}{T^2}\right) = O\left(\lambda_j^{-2d_0} \left(\frac{k_T}{T}\right)^2\right)$$

by assumption A.3, where $g_x(\lambda) = g_v(\lambda)$ if $d_0 < 1/2$ and $g_x(\lambda) = g_v(\lambda)|2\lambda^{-1}\sin(\lambda/2)|^{-2}$ if $d_0 \ge 1/2$. Now the first summand in (A.4) is equal to

$$\lambda_j^{-2d_0} \sum_{i\in\Delta_T} p_i |\lambda_j+\lambda_i|^{2d_0} (I_{j+i} - |\alpha_{j+i}|^2 I_{\varepsilon,j+i}) + \lambda_j^{-2d_0} \sum_{i\in\Delta_T} p_i |\lambda_j+\lambda_i|^{2d_0} (|\alpha_{j+i}|^2 I_{\varepsilon,j+i} - f_{x,j+i}) \tag{A.5}$$

where $\alpha_{j+i} = \alpha(\lambda_{j+i})$ for $\alpha(\lambda) = \sum_{k=0}^{\infty} b_k e^{ik\lambda}$ in the stationary case and $\alpha(\lambda) = (1 - e^{i\lambda})^{-1} \sum_{k=0}^{\infty} b_k e^{ik\lambda}$ if $d_0 \geq 1/2$. Using (3.17) in Robinson (1995) and formula (A1) in Velasco (1999) for the nonstationary case we get that the first term in (A.5) is

$$
\begin{aligned}
& O_p\left(\lambda_j^{-2d_0} \sum_{i\in\Delta_T} p_i |\lambda_j + \lambda_i|^{2d_0} f_{i+j} \frac{\log^{1/2}|j+i|}{|j+i|^{\alpha}}\right) \\
= \; & O_p\left(\lambda_j^{-2d_0} \sum_{i\in\Delta_T} p_i \frac{\log^{1/2}|j+i|}{|j+i|^{\alpha}}\right) \\
= \; & O_p\left(\lambda_j^{-2d_0} (\sum_{i\in\Delta_T} p_i^2)^{1/2} \left(\sum_{i\in\Delta_T} \frac{\log|j+i|}{|j+i|^{2\alpha}}\right)^{1/2}\right) \\
= \; & O_p\left(\lambda_j^{-2d_0} k_T^{-1/2} \log^{1/2}|j+k_T| \left(\sum_{i\in\Delta_T} |j+i|^{-2\alpha}\right)^{1/2}\right) \qquad \text{(A.6)} \\
= \; & O_p\left(\lambda_j^{-2d_0} \left(\frac{\log|j+k_T|}{k_T^{1/2}} \mathbb{I}_{(\alpha=1/2)} + \frac{\log^{1/2}|j+k_T|}{k_T^{1/2}} k_T^{1/2-\alpha} \left|\frac{j}{k_T}+1\right|^{1/2-\alpha} \mathbb{I}_{(\alpha<1/2)}\right)\right) \\
= \; & O_p\left(\lambda_j^{-2d_0} \left(\frac{\log|j+k_T|}{k_T^{1/2}} \mathbb{I}_{(\alpha=1/2)} + \frac{\log^{1/2}|j+k_T|}{k_T^{\alpha}} \mathbb{I}_{(\alpha<1/2)}\right)\right)
\end{aligned}
$$

if $j/k_T \to 0$, and in any other case, using Lemma 1 below, (A.6) is

$$
\begin{aligned}
& O_p\left(\lambda_j^{-2d_0} \left(\frac{\log|j+k_T|}{k_T^{1/2}} \mathbb{I}_{(\alpha=1/2)} + \frac{\log^{1/2}|j+k_T|}{k_T^{1/2}} k_T^{1/2} j^{-\alpha} \mathbb{I}_{(\alpha<1/2)}\right)\right) \\
= \; & O_p\left(\lambda_j^{-2d_0} \left(\frac{\log|j+k_T|}{k_T^{1/2}} \mathbb{I}_{(\alpha=1/2)} + \frac{\log^{1/2}|j+k_T|}{j^{\alpha}} \mathbb{I}_{(\alpha<1/2)}\right)\right)
\end{aligned}
$$

such that the first term in (A.5) is

$$
O_p\left(\lambda_j^{-2d_0} \left(\frac{\log|j+k_T|}{k_T^{1/2}} \mathbb{I}_{(\alpha=1/2)} + \frac{\log^{1/2}|j+k_T|}{\overline{jk_T}^{\alpha}} \mathbb{I}_{(\alpha<1/2)}\right)\right)
$$

where $\overline{jk_T} = k_T$ if $j/k_T \to 0$ and $\overline{jk_T} = j$ otherwise.

Now the second term in (A.5) is equal to

$$
\lambda_j^{-2d_0} \sum_{i\in\Delta_T} p_i g_x(\lambda_j + \lambda_i)(2\pi I_{\varepsilon,j+i} - 1)
$$

which has mean zero and variance

$$
\begin{aligned}
& \lambda_j^{-4d_0} \sum_{i\in\Delta_T} p_i^2 g_x^2(\lambda_j+\lambda_i) 4\pi^2 var(I_{\varepsilon,j+i}) \\
+ \; & \lambda_j^{-4d_0} \sum_{i\in\Delta_T} \sum_{l\neq i} p_i p_l g_x(\lambda_j+\lambda_i) g_x(\lambda_j+\lambda_l) 4\pi^2 cov(I_{\varepsilon,j+i}, I_{\varepsilon,j+l}) \\
= \; & O\left(\lambda_j^{-4d_0} \sum_{i\in\Delta_T} p_i^2\right) + O\left(\lambda_j^{-4d_0} \frac{1}{T} \sum_{i\in\Delta_T} \sum_{l\neq i} p_i p_l\right) \\
= \; & O\left(\lambda_j^{-4d_0} \left[\frac{1}{k_T} + \frac{1}{T}\right]\right)
\end{aligned}
$$

by for example, Proposition 10.3.2 in Brockwell and Davis (2006).

Finally, in order to get a bound for (A.3) consider the case $k_T > j$ (when $k_T \leq j$ the analysis is similar but there is not need to split Δ_T in two sets, before and after $-j$). Then (A.3) is equal to

$$\left\{ \sum_{i=-k_T+j}^{-1} + \sum_{i=1}^{k_T+j} \right\} p_{i-j} I_i \left|\frac{i}{j}\right|^{2d_0} \left(\left|\frac{i}{j}\right|^{2(\hat{d}-d_0)} - 1 \right) = O_p(f_j A_T \log T)$$

because

$$\left|\frac{i}{j}\right|^{2d_0} - 1 \leq constant \times (\hat{d}-d_0) \log\left|\frac{i}{j}\right|.$$

$\blacksquare$

Proof of Theorem 2: Let $\bar{v}_j^* = G2\pi \bar{I}_{\varepsilon j}$ where $G = g_v(0)$ and $\bar{I}_{\varepsilon j}$ takes values from the set $\{I_{\varepsilon j+i}\}_{i\in\Delta_T}$ with probability p_i. For simplicity of notation denote $d_2(\hat{v}_j^*, v_j^0) = d_2[\mathcal{L}(\hat{v}_j^*|x_1,...,x_T), \mathcal{L}(v_j^0)]$. Then

$$d_2^2(\hat{v}_j^*, v_j^0) \leq d_2^2(\hat{v}_j^*, \bar{v}_j^*) + d_2^2(\bar{v}_j^*, v_j^0)$$

uniformly in j. Now

$$
\begin{aligned}
d_2^2(\hat{v}_j^*, \bar{v}_j^*) \; & \leq \; E^*|\hat{v}_j^* - \bar{v}_j^*|^2 \\
& = \; G^2 \sum_{i\in\Delta_T} p_i \left| \frac{I_{j+i}}{G\lambda_{j+i}^{-2\hat{d}}} - 2\pi I_{\varepsilon j+i} \right|^2 .
\end{aligned}
$$

Taking into account that $E\left|\frac{I_{j+i}}{G\lambda_{j+i}^{-2d_0}}\right| < constant$ for $j = 1, ..., m$ (see Robinson, 1995, formula (3.16) for the stationary case and Velasco, 1999, formula (A1) for nonstationary series) and that $(\hat{d}-d_0) = o_p(\log^{-1} T)$ then

$$\frac{I_{j+i}}{G\lambda_{j+i}^{-2\hat{d}}} - 2\pi I_{\varepsilon j+i} = \lambda_{j+i}^{2(\hat{d}-d_0)} \frac{I_{j+i}}{G\lambda_{j+i}^{-2d_0}} - 2\pi I_{\varepsilon j+i} = \frac{I_{j+i}}{G\lambda_{j+i}^{-2d_0}} - 2\pi I_{\varepsilon j+i} + o_p(1).$$

Now

$$\frac{I_{j+i}}{G\lambda_{j+i}^{-2d_0}} - 2\pi I_{\varepsilon j+i} = \left(1 - \frac{G\lambda_{j+i}^{-2d_0}}{f_{x,j+i}}\right) \frac{I_{j+i}}{G\lambda_{j+i}^{-2d_0}} + \frac{I_{j+i}}{f_{x,j+i}} - 2\pi I_{\varepsilon j+i} \tag{A.7}$$

and by assumptions A.1, A.2 and formulae (3.17) and (A2) in Robinson (1995) and Velasco (1995) respectively

$$\begin{aligned} 1 - \frac{G\lambda_j^{-2d_0}}{f_{x,j}} &= O(\lambda_j^2) \\ \left| \frac{I_j}{f_{x,j}} - 2\pi I_{\varepsilon j} \right| &= O_p\left(j^{-\alpha} \log^{1/2} j\right) \end{aligned}$$

such that using Cauchy–Schwarz inequality and assumption A.3,

$$\begin{aligned} d_2^2(\hat{v}_j^*, \bar{v}_j^*) &= O_p\left(\sum_{i \in \Delta_T} p_i \left[\lambda_{j+i}^4 + |j+i|^{-2\alpha} \log|j+i|\right]\right) + o_p(1) \\ &= o_p(1) + O_p\left(k_T^{-1/2}\left[\frac{(j+k_T)^4}{T^4} k_T^{1/2} + j^{-2\alpha} k_T^{1/2} \log|j+k_T|\right]\right) \\ &= o_p(1) + O_p\left(j^{-2\alpha} \log|j+k_T|\right) \end{aligned}$$

where the $o_p(1)$ term holds uniformly in $j = 1, ..., m$ and the $O_p()$ term comes from the bound in Lemma 1.

Next

$$d_2(\bar{v}_j^*, v_j^0) \leq d_2(\bar{v}_j^*, \bar{v}_j^+) + d_2(\bar{v}_j^+, \bar{v}_j) + d_2(\bar{v}_j, v_j^0)$$

for $\bar{v}_j = G2\pi I_{\varepsilon j}$ and $\bar{v}_j^+ = GK_j$ where K_j are independent standard exponentially distributed variables. By Lemma A.1 in Paparoditis and Politis (1999) and using the convergence of $I_{\varepsilon j}$ to the exponential distribution and Lemma 8.3 in Bickel and Freedman (1981), $d_2^2(\bar{v}_j^*, \bar{v}_j^+) \to 0$ as $T \to \infty$, noting also that $E(2\pi I_{\varepsilon j}) \to 1$ and $Var(2\pi I_{\varepsilon j}) \to 1$ uniformly in $j = 1, ..., m$. Now $d_2^2(\bar{v}_j^+, \bar{v}_j) \leq GE|2\pi I_{\varepsilon j} - K_j|^2 \to 0$ by the uniform convergence of $I_{\varepsilon j}$ to a random variable with exponential distribution. Finally,

$$d_2^2(\bar{v}_j, v_j^0) \leq E\left|\frac{I_j}{G\lambda_j^{-2\hat{d}}} - 2\pi I_{\varepsilon j}\right|^2$$

and using (A.7), assumptions A.2 and A.4, formula (3.16) in Robinson (1995), formula (A1) in Velasco (1999),

$$d_2^2(\bar{v}_j, v_j^0) = o(1) + O\left(\frac{\log j}{j^{2\alpha}}\right)$$

because using the results in Robinson (1995), pages 1648-1651 (for the stationary case) and the proof of Lemma 1 in Velasco (1999) (for nonstationary series), $E|I_j/f_{x,j} - 2\pi I_{\varepsilon j}|^2 = O(j^{-2a} \log j)$.

■

Proof of Corollary 1: Let $\bar{I}_j^* = \lambda_j^{-2d_0}\hat{v}_j^*$. Then

$$d_2^2(I_j^*, I_j) \leq d_2^2(I_j^*, \bar{I}_j^*) + d_2^2(\bar{I}_j^*, I_j)$$

Now

$$\begin{aligned} d_2^2(I_j^*, \bar{I}_j^*) &\leq E^* \left|(\lambda_j^{-2\hat{d}} - \lambda_j^{-2d_0})\hat{v}_j^*\right|^2 = \sum_{i\in\Delta_T} p_i \left|(\lambda_j^{-2\hat{d}} - \lambda_j^{-2d_0})\hat{v}_{j+i}\right|^2 \\ &= o_p\left(\lambda_j^{-4d_0}\right) \end{aligned}$$

because $E\left|\frac{I_{j+i}}{G\lambda_{j+i}^{-2d_0}}\right| < constant$ for $j = 1, ..., m$ and $\lambda_j^{-2\hat{d}} - \lambda_j^{-2d_0} = O_p(\lambda_j^{-2d_0}\log T(\hat{d} - d_0)) = o_p(\lambda_j^{-2d_0})$.

Finally, by Theorem 2

$$d_2^2(\bar{I}_j^*, I_j) = \lambda_j^{-4d_0} d_2^2(\hat{v}_j^*, v_j^0) = o_p(\lambda_j^{-4d_0}) + O_p\left(\lambda_j^{-4d_0}\frac{\log(k_T + j)}{j^{2\alpha}}\right)$$

■

Proof of Theorem 3: For simplicity of notation denote $d_2(\Phi_m^*, \Phi_m) = d_2[\mathcal{L}(\Phi_m^*|x_1, ..., x_T), \mathcal{L}(\Phi_m)]$. Consider also the following statistics:

$$\begin{aligned} \hat{\Phi}_m &= \sum_{j=1}^m \psi_j \lambda_j^{-2d_0}\hat{v}_j^* \\ \bar{\Phi}_m &= \sum_{j=1}^m \psi_j \lambda_j^{-2d_0} 2\pi G I_{\varepsilon j} = \sum_{j=1}^m \psi_j \lambda_j^{-2d_0}\bar{v}_j \\ \Phi_m^+ &= \sum_{j=1}^m \psi_j \lambda_j^{-2d_0} G K_j = \sum_{j=1}^m \psi_j \lambda_j^{-2d_0}\bar{v}_j^+ \\ \hat{\Phi}_m^c &= \sum_{j=1}^m \psi_j \lambda_j^{-2d_0} G \frac{\hat{v}_j^*}{\bar{v}_j^c} \text{ where } \bar{v}_j^c = \sum_{i\in\Delta_T} p_i \hat{v}_{j+i}. \end{aligned}$$

Then

$$d_2(\Phi_m^*, \Phi_m) \leq d_2(\Phi_m^*, \hat{\Phi}_m) + d_2(\hat{\Phi}_m, \hat{\Phi}_m^c) + d_2(\hat{\Phi}_m^c, \Phi_m^+) + d_2(\Phi_m^+, \bar{\Phi}_m) + d_2(\bar{\Phi}_m, \Phi_m)$$

The first term in the right hand side $d_2(\Phi_m^*, \hat{\Phi}_m)$ is bounded by

$$\begin{aligned} &\left\{E^*\left[\sum \psi_j \hat{v}_j^* \lambda_j^{-2d_0}(\lambda_j^{2(d_0-\hat{d})} - 1)\right]^2\right\}^{1/2} \\ \leq\ & \sum |\psi_j| \lambda_j^{-2d_0} |\lambda_j^{2(d_0-\hat{d})} - 1| \left(E^*|\hat{v}_j^*|^2\right)^{1/2} \\ =\ & O_p\left(A_T \log T \sum |\psi_j| \lambda_j^{-2d_0}\right) \end{aligned}$$

using Minkowski inequality and the facts that $(d_0 - \hat{d}) = O_p(A_T)$ and $E^*|\hat{v}_j^*|^2 = \sum_{i\in\Delta_T} p_i |\hat{v}_{j+i}|^2 = O_p(1)$.

Next, $d_2(\hat{\Phi}_m, \hat{\Phi}_m^c)$ is bounded by

$$\begin{aligned} & \left\{ E^* \left[\sum \psi_j \lambda_j^{-2d_0} \hat{v}_j^* \left(1 - \frac{G}{\bar{v}_j^c}\right) \right]^2 \right\}^{1/2} \\ \leq \; & \sum |\psi_j| \lambda_j^{-2d_0} \left| 1 - \frac{G}{\bar{v}_j^c} \right| \left(E^* |\hat{v}_j^*|^2\right)^{1/2} \\ = \; & O_p\left(\sum |\psi_j| \lambda_j^{-2d_0} B(j, k_T, T) \right). \end{aligned}$$

The inequality is based on Minkowsky inequality and the bound in probability is because $E^*|\hat{v}_j^*|^2 = O_p(1)$ and

$$\begin{aligned} \bar{v}_j^c &= \lambda_j^{2(\hat{d}-d_0)} \lambda_j^{2d_0} E^*(I_j^*) = \lambda_j^{2d_0} E^*(I_j^*)(1 + O_p(A_T \log T)) \\ &= G(1 + O_p(B(j, k_T, T))) \end{aligned}$$

by Theorem 1 and assumption A.2.

Regarding the third term, note that

$$d_2^2(\hat{\Phi}_m^c, \Phi_m^+) \leq \sum \psi_j^2 \lambda_j^{-4d_0} G^2 d_2^2 \left(\frac{\hat{v}_j^*}{\bar{v}_j^c}, K_j \right)$$

by Lemma 8.7 in Bickel and Freedman (1981) because $\left(\frac{\hat{v}_j^*}{\bar{v}_j^c}, K_j\right)$ are independent for different j and

$$E^* \left(\frac{\hat{v}_j^*}{\bar{v}_j^c} \right) = E(K_j) = 1.$$

Now

$$d_2^2 \left(\frac{\hat{v}_j^*}{\bar{v}_j^c}, K_j \right) \leq d_2^2 \left(\frac{\hat{v}_j^*}{\bar{v}_j^c}, \frac{\hat{v}_j^*}{G} \right) + d_2^2 \left(\frac{\hat{v}_j^*}{G}, K_j \right)$$

such that

$$d_2^2 \left(\frac{\hat{v}_j^*}{\bar{v}_j^c}, \frac{\hat{v}_j^*}{G} \right) \leq E^* \left(\frac{\hat{v}_j^*}{\bar{v}_j^c} - \frac{\hat{v}_j^*}{G} \right)^2 = \sum_{i\in\Delta_T} p_i \hat{v}_{j+i}^2 \left(\frac{1}{\bar{v}_j^c} - \frac{1}{G} \right)^2 = O_p(B^2(j, k_T, T)),$$

and as in the proof of Theorem (2)

$$d_2^2 \left(\frac{\hat{v}_j^*}{G}, K_j \right) = G^2 d_2^2 \left(\hat{v}_j^*, \bar{v}_j^+ \right) = o_p(1) + O_p\left(\frac{\log(j + k_T)}{j^{2\alpha}} \right).$$

Next, using again Lemma 8.7 in Bickel and Freedman (1981), $d_2^2(\Phi_m^+, \bar{\Phi}_m)$ is bounded by

$$\sum \psi_j^2 \lambda_j^{-4d_0} d_2^2(K_j, 2\pi I_{\varepsilon j}) = o(\sum \psi_j^2 \lambda_j^{-4d_0})$$

by the uniform convergence of $2\pi I_{\varepsilon j}$ to a standard exponential, $E(2\pi I_{\varepsilon j}) = 1$ and $Var(2\pi I_{\varepsilon j}) = 1$.

Finally $d_2^2(\bar{\Phi}_m, \Phi_m)$ is bounded by

$$
\begin{aligned}
& E\left\{\sum \psi_j \lambda_j^{-2d_0} G\left(\frac{I_j}{G\lambda_j^{-2d_0}} - 2\pi I_{\varepsilon j}\right)\right\}^2 \\
= & \sum \psi_j^2 \lambda_j^{-4d_0} G^2 E\left(\frac{I_j}{G\lambda_j^{-2d_0}} - 2\pi I_{\varepsilon j}\right)^2 \\
+ & 2\sum_j \sum_{l>j} \psi_j \psi_l \lambda_j^{-2d_0} \lambda_l^{-2d_0} G^2 E\left(\frac{I_j}{G\lambda_j^{-2d_0}} - 2\pi I_{\varepsilon j}\right)\left(\frac{I_l}{G\lambda_l^{-2d_0}} - 2\pi I_{\varepsilon l}\right) \\
= & O\left(\sum \psi_j^2 \lambda_j^{-4d_0} \frac{\log j}{j^{2\alpha}}\right) \\
+ & O\left(\sum_j \sum_{l>j} \psi_j \psi_l \lambda_j^{-2d_0} \lambda_l^{-2d_0} \left[\frac{\log^2 m}{j^{4\alpha}} + \frac{\log^{1/2} m}{j^{2\alpha} l^{\alpha}}\right]\right)
\end{aligned}
$$

using the results in Robinson (1995), pages 1648-1651, for the stationary case and page 118 in Velasco (1999) for the non-stationary one.

■

Proof of Corollary 3: The proof is a direct application of the results of Theorem 3 noting that

$$
\sum \psi_j^2 \lambda_j^{-4d_0} = \frac{4}{m g_v(0)} \sum v_j^2 \to \frac{4}{m g_v(0)}
$$

because $\sum v_j^2 = 1 + o(1)$. The only modification required is in the bound of $d_2^2(\bar{\Phi}_m, \Phi_m)$, which can be strengthened in this case as follows. Note that $d_2^2(\bar{\Phi}_m, \Phi_m)$ is bounded by

$$
\begin{aligned}
& E\left\{\sum \psi_j \lambda_j^{-2d_0} G\left(\frac{I_j}{G\lambda_j^{-2d_0}} - 2\pi I_{\varepsilon j}\right)\right\}^2 \\
= & E\left\{\sum \psi_j \lambda_j^{-2d_0} G\left(A_{1j} + A_{2j}\right)\right\}^2 \text{ for } A_{1j} = \frac{I_j}{G\lambda_j^{-2d_0}} - \frac{I_j}{f_j} \text{ and } A_{2j} = \frac{I_j}{f_j} - 2\pi I_{\varepsilon j}
\end{aligned}
$$

Now $E\left(\sum \psi_j \lambda_j^{-2d_0} G A_{1j}\right)^2$ is

$$
\begin{aligned}
& \sum \psi_j^2 \lambda_j^{-4d_0} G^2 E\left(A_{1j}^2\right) + 2\sum_j \sum_{l>j} \psi_j \psi_l \lambda_j^{-2d_0} \lambda_l^{-2d_0} G^2 E\left(A_{1j} A_{1l}\right) \\
= & O\left(\frac{1}{m}\left[\sum v_j^2 E\left(A_{1j}^2\right) + \sum_j \sum_{l>j} v_j v_l E\left(A_{1j} A_{1l}\right)\right]\right) \\
= & O\left(\frac{m^4}{T^4} \log^2 m + \frac{m^5}{T^4} \log^2 m\right) = o(1)
\end{aligned}
$$

under assumption A.7, using

$$E\left(A_{1j}^2\right) = \left(1 - \frac{G\lambda_j^{-2d_0}}{f_j}\right)^2 E\left|\frac{I_j}{G\lambda_j^{-2d_0}}\right|^2 = O(\lambda_j^4)$$

and

$$E\left(A_{1j}A_{1l}\right) = \left(1 - \frac{G\lambda_j^{-2d_0}}{f_j}\right)\left(1 - \frac{G\lambda_l^{-2d_0}}{f_l}\right) E\left(\frac{I_j}{G\lambda_j^{-2d_0}}\frac{I_l}{G\lambda_l^{-2d_0}}\right) = O(\lambda_j^2\lambda_l^2)$$

because

$$E\left(\frac{I_j}{G\lambda_j^{-2d_0}}\frac{I_l}{G\lambda_l^{-2d_0}}\right) = \frac{f_j f_l}{G^2\lambda_j^{-2d_0}\lambda_l^{-2d_0}} E\left(\frac{I_j}{f_j}\frac{I_l}{f_l}\right) = O(1)$$

using the results in Robinson (1995, pages 1648-51) and the details in the proof of Lemma 1 in Velasco (1999).

Next

$$E\left(\sum \psi_j \lambda_j^{-2d_0} G A_{2j}\right)^2 = O\left(\frac{1}{m}E\left[\sum_{j=1}^{\sqrt{m}} v_j A_{2j}\right]^2\right) + O\left(\frac{1}{m}E\left[\sum_{j=\sqrt{m}+1}^{m} v_j A_{2j}\right]^2\right). \quad \text{(A.8)}$$

Using Minkowski inequality, the first bound in (A.8) is

$$O\left(\frac{1}{m}\left[\sum_{j=1}^{\sqrt{m}} |v_j| E\{A_{2j}^2\}^{1/2}\right]^2\right) = O\left(\frac{1}{m}\log^2 m\left[\sum_{j=1}^{\sqrt{m}} \frac{\log^{1/2} j}{j^{\alpha}}\right]^2\right) = O\left(\frac{1}{m}m^{1-\alpha}\log^3 m\right) = o(1)$$

because $E(A_{2j}^2) = O(j^{-2al}\log j)$ using the results in Robinson (1995), pages 1648-1651, for the stationary case and page 118 in Velasco (1999) for the non-stationary one. Finally, using summation by parts the second bound in (A.8) is

$$O\left(\frac{1}{m}E\left[\sum_{r=\sqrt{m}+1}^{m-1}(v_r - v_{r+1})\sum_{i=\sqrt{m}+1}^{r} A_{2i}\right]^2\right) + O\left(\frac{v_m^2}{m}E\left[\sum_{i=\sqrt{m}+1}^{m} A_{2i}\right]^2\right). \quad \text{(A.9)}$$

Because $|v_{r+1} - v_r| = |\log(1 + r^{-1})| \le r^{-1}$ and using Minkowski inequality the first bound in (A.9) is

$$O\left(\frac{1}{m}\left\{\sum_{r=\sqrt{m}+1}^{m-1}|v_r - v_{r+1}|\left[E\left|\sum_{i=\sqrt{m}+1}^{r} A_{2i}\right|^2\right]^{1/2}\right\}^2\right)$$

$$O\left(\frac{1}{m}\left\{\sum_{r=\sqrt{m}+1}^{m-1}\frac{1}{r}\left[E\left|\sum_{i=\sqrt{m}+1}^{r} A_{2i}\right|^2\right]^{1/2}\right\}^2\right)$$

$$O\left(\frac{1}{m}\left\{\sum_{r=\sqrt{m}+1}^{m-1}\frac{1}{r}\left[a+b\right]^{1/2}\right\}^2\right) \quad \text{(A.10)}$$

where a and b are as defined in Robinson (1995, page 1648) and Velasco (1999, page 118). Using the results in both papers, (A.10) is

$$
\begin{aligned}
& O\left(\frac{1}{m}\left\{\sum_{r=\sqrt{m}+1}^{m-1}\frac{1}{r}\left[\frac{\log^2 r}{r^{2(2\alpha-1)}}+r^{1/2}\log r+\frac{r\log^2 r}{m^{2\alpha-1/2}}+\frac{\log^{5/2}T}{Tr^{\alpha-2}}\mathbb{I}_{(\alpha<1/2)}+\frac{\log r}{T^{1/2}r^{2\alpha-2}}\right]^{1/2}\right\}^2\right) \\
= \;& O\left(\frac{\log^2 m}{m^{4\alpha-1}}+\frac{\log m}{m^{1/2}}+\frac{\log^2 m}{m^{2\alpha-1/2}}+\frac{\log m}{m^{1/2}}+\frac{\log^{5/2}T}{Tm^{\alpha-1}}+\frac{\log m}{T^{1/2}m^{2\alpha-1}}\right) \\
= \;& o(1)
\end{aligned}
$$

under assumption A.7 if $d_0 < 3/4$ such that $\alpha > 1/4$. Finally the second bound in (A.9)

$$
\begin{aligned}
& O\left(\frac{\log^2 m}{m}\left[\frac{\log^2 m}{m^{2(2\alpha-1)}}+m^{1/2}\log m+\frac{m\log^2 m}{m^{2\alpha-1/2}}+\frac{\log^{5/2}T}{Tm^{\alpha-2}}\mathbb{I}_{(\alpha<1/2)}+\frac{\log m}{T^{1/2}m^{2\alpha-2}}\right]\right) \\
= \;& \left(\frac{\log^4 m}{m^{2(2\alpha-1/2)}}+\frac{\log^3 m}{m^{1/2}}+\frac{\log^4 m}{m^{2\alpha-1/2}}+\frac{\log^{5/2}T\log^2 m}{Tm^{\alpha-1}}\mathbb{I}_{(\alpha<1/2)}+\frac{\log^3 m}{T^{1/2}m^{2\alpha-1}}\right) \\
= \;& o(1)
\end{aligned}
$$

■

Lemma 1 *Let $\Delta_T = \{0, \pm 1, , ..., \pm k_T\}\backslash\{-j\}$. If $\beta > 0$ then*

$$\sum_{i\in\Delta_T}|j+i|^{-\beta} = O(\log|k_T+j|\mathbb{I}_{(\beta\geq 1)}+k_Tj^{-\beta}\mathbb{I}_{(\beta<1)})$$

.

Proof: Consider first the case $k_T < j$ such that $\Delta_T = \{-k_T, ..., k_T\}$. Then, if $k_T/j \to 0$ as $T \to \infty$,

$$\sum_{i=-k_T}^{k_T}|j+i|^{-\beta} = j^{-\beta}\sum_{i=-k_T}^{k_T}\left|1+\frac{i}{j}\right|^{-\beta} = j^{-\beta}O(k_T).$$

If $k_T = bj$ for $b < 1$, then

$$\sum_{i=-k_T}^{k_T}|j+i|^{-\beta} \leq 3k_T\,|j-k_T|^{-\beta} = O(j^{-\beta}k_T);$$

and if $j - k_T = \textit{constant} > 0$ then

$$\sum_{i=-k_T}^{k_T}|j+i|^{-\beta} \;=\; \sum_{i=j-k_T}^{j+k_T}|i|^{-\beta} = O\left(\log|j+k_T|\mathbb{I}_{(\beta\geq 1)}+k_Tj^{-\beta}\mathbb{I}_{(\beta<1)}\right).$$

Next, if $k_T = j$ then $\Delta_T = \{-k_T+1, ..., k_T\}$ and

$$\sum_{i=-k_T+1}^{k_T}|j+i|^{-\beta} = \sum_{i=1}^{j+k_T}|i|^{-\beta} = O\left(\log|j+k_T|\mathbb{I}_{(\beta\geq 1)}+k_Tj^{-\beta}\mathbb{I}_{(\beta<1)}\right).$$

Finally, consider the case $k_T > j$ such that $\Delta_T = \{-k_T, ..., -j-1\} \cup \{-j+1, ..., k_T\}$. When $j = bk_T$ for $b < 1$ or $k_T - j = constant > 0$, as before

$$\sum_{i=-j+1}^{k_T} |j+i|^{-\beta} = \sum_{i=1}^{k_T+j} i^{-\beta} = O\left(\log|j+k_T|\mathbb{I}_{(\beta\geq 1)} + k_T j^{-\beta}\mathbb{I}_{(\beta<1)}\right),$$

because for $\beta < 1$ $(k_T + j)^{1-\beta} = k_T^{1-\beta}(1 + j/k_T)^{1-\beta} = O(k_T^{1-\beta}) = O(k_T j^{-\beta}[k/j]^{-\beta})$. Similarly

$$\begin{aligned}\sum_{i=-k_T}^{-j-1} |j+i|^{-\beta} &= \sum_{i=1}^{k_T-j} i^{-\beta} = O\left(\log|k_T|\mathbb{I}_{(\beta\geq 1)} + (k_T-j)^{1-\beta}\mathbb{I}_{(\beta<1)}\right) \\ &= O\left(\log|k_T|\mathbb{I}_{(\beta\geq 1)} + k_T j^{-\beta}\mathbb{I}_{(\beta<1)}\right).\end{aligned}$$

Now if $j/k_T \to 0$,

$$\begin{aligned}\sum_{i=-j+1}^{k_T} |j+i|^{-\beta} &= \sum_{i=1}^{j+k_T} i^{-\beta} = O\left(\log|k_T+j|\mathbb{I}_{(\beta\geq 1)} + k_T^{1-\beta}\left|1+\frac{j}{k_T}\right|^{1-\beta}\mathbb{I}_{(\beta<1)}\right) \\ &= O\left(\log|k_T+j|\mathbb{I}_{(\beta\geq 1)} + k_T j^{-\beta}\mathbb{I}_{(\beta<1)}\right), \\ \sum_{i=-k_T}^{-j-1} |j+i|^{-\beta} &= \sum_{i=1}^{k_T-j} i^{-\beta} = O\left(\log k_T\mathbb{I}_{(\beta\geq 1)} + k_T^{1-\beta}\left|1-\frac{j}{k_T}\right|^{1-\beta}\mathbb{I}_{(\beta<1)}\right) \\ &= O\left(\log k_T\mathbb{I}_{(\beta\geq 1)} + k_T j^{-\beta}\mathbb{I}_{(\beta<1)}\right).\end{aligned}$$

■

B Appendix: Monte Carlo results in ARFIMA(1,d,0)

We complement the results in the Monte Carlo in Section 5 with an analogous analysis applied to the ARFIMA model:

$$(1-0.6L)(1-L)^{d_0}X_t = u_t, \quad t = 1, 2, ..T, \tag{B.1}$$

where the u_t are standard normal. For values of $d_0 \geq 1/2$ the series is obtained as described in Assumption A.1 by integration of an $ARFIMA(1, d_0 - 1, 0)$ process. The number of replications is 1000 and the memory parameters considered are $d_0 = -04$, 0.4 and 0.7.

Table 5: Ratio of RMSD: ARFIMA(1,d_0,0), $\phi = 0.6$

	$d_0 = -0.4$, $T = 64$								
	$m = 3$			$m = 8$			$m = 15$		
	$m_1 = 3$	$m_1 = 8$	$m_1 = 15$	$m_1 = 3$	$m_1 = 8$	$m_1 = 15$	$m_1 = 3$	$m_1 = 8$	$m_1 = 15$
PFSBS	**0.596**	0.887	1.043	1.133	1.163	0.879	0.971	1.030	1.030
PFSDDB	0.704	0.970	1.189	1.334	1.302	0.749	**0.913**	1.046	1.090
FDLB ($k_T = 3$)	2.259	0.942	0.996	1.546	1.543	1.094	1.039	1.141	1.114
FDLB ($k_T = 5$)	1.895	0.861	1.048	1.448	1.438	0.799	0.998	1.125	1.098
FDLB ($k_T = 10$)	1.292	0.838	1.085	1.552	1.383	**0.660**	0.948	1.071	1.074
FDLB ($k_T = 25$)	0.983	0.985	1.159	1.587	1.380	0.746	0.963	1.045	1.089
	$d_0 = -0.4$, $T = 128$								
	$m = 5$			$m = 15$			$m = 30$		
	$m_1 = 5$	$m_1 = 15$	$m_1 = 30$	$m_1 = 5$	$m_1 = 15$	$m_1 = 30$	$m_1 = 5$	$m_1 = 15$	$m_1 = 30$
PFSBS	0.242	0.789	1.004	1.599	1.533	0.936	1.008	1.025	1.010
PFSDDB	**0.207**	0.805	1.080	1.714	1.861	0.744	**0.958**	1.035	1.034
FDLB ($k_T = 5$)	0.673	0.510	0.739	2.102	2.012	1.357	1.012	1.042	1.033
FDLB ($k_T = 10$)	0.505	0.611	0.941	1.878	1.850	0.821	0.988	1.042	1.027
FDLB ($k_T = 20$)	0.513	0.713	1.012	1.995	1.869	**0.699**	0.966	1.035	1.022
FDLB ($k_T = 50$)	0.797	0.933	1.075	1.998	1.878	0.788	0.965	1.032	1.034
	$d_0 = 0.4$, $T = 64$								
	$m = 3$			$m = 8$			$m = 15$		
	$m_1 = 3$	$m_1 = 8$	$m_1 = 15$	$m_1 = 3$	$m_1 = 8$	$m_1 = 15$	$m_1 = 3$	$m_1 = 8$	$m_1 = 15$
PFSBS	0.572	0.885	0.977	0.694	0.715	0.520	0.920	0.944	0.940
PFSDDB	0.487	0.913	1.064	0.807	0.874	0.450	**0.878**	1.004	1.037
FDLB ($k_T = 3$)	1.795	**0.460**	0.484	1.014	0.998	0.704	0.987	1.089	1.058
FDLB ($k_T = 5$)	1.304	0.480	0.670	0.958	0.935	0.529	0.949	1.073	1.042
FDLB ($k_T = 10$)	0.692	0.629	0.787	1.027	0.901	**0.449**	0.917	1.020	1.025
FDLB ($k_T = 25$)	0.716	0.950	0.970	1.059	0.919	0.506	0.940	1.003	1.054
	$d_0 = 0.4$, $T = 128$								
	$m = 5$			$m = 15$			$m = 30$		
	$m_1 = 5$	$m_1 = 15$	$m_1 = 30$	$m_1 = 5$	$m_1 = 15$	$m_1 = 30$	$m_1 = 5$	$m_1 = 15$	$m_1 = 30$
PFSBS	0.171	0.978	1.310	0.763	0.693	0.395	0.965	0.991	0.972
PFSDDB	**0.132**	0.993	1.376	0.858	0.893	**0.339**	**0.921**	1.013	1.014
FDLB ($k_T = 5$)	0.687	0.639	0.952	1.072	1.011	0.683	0.990	1.027	1.017
FDLB ($k_T = 10$)	0.313	0.776	1.191	0.982	0.932	0.425	0.964	1.028	1.012
FDLB ($k_T = 20$)	0.427	0.891	1.269	1.059	0.953	0.374	0.947	1.020	1.008
FDLB ($k_T = 50$)	0.854	1.078	1.310	1.061	0.980	0.416	0.949	1.016	1.025
	$d_0 = 0.7$, $T = 64$								
	$m = 3$			$m = 8$			$m = 15$		
	$m_1 = 3$	$m_1 = 8$	$m_1 = 15$	$m_1 = 3$	$m_1 = 8$	$m_1 = 15$	$m_1 = 3$	$m_1 = 8$	$m_1 = 15$
PFSBS	0.694	0.800	0.886	0.918	0.868	0.728	1.026	1.018	1.013
PFSDDB	0.681	0.846	0.954	0.946	0.963	0.687	1.028	1.097	1.109
FDLB ($k_T = 3$)	2.149	0.589	0.597	1.033	1.007	0.779	1.058	1.169	1.134
FDLB ($k_T = 5$)	1.563	**0.584**	0.724	0.970	0.958	0.628	1.012	1.150	1.114
FDLB ($k_T = 10$)	0.885	0.671	0.811	1.044	0.943	**0.570**	**0.973**	1.091	1.108
FDLB ($k_T = 25$)	0.788	0.877	0.899	1.064	0.973	0.618	1.005	1.053	1.157
	$d_0 = 0.7$, $T = 128$								
	$m = 5$			$m = 15$			$m = 30$		
	$m_1 = 5$	$m_1 = 15$	$m_1 = 30$	$m_1 = 5$	$m_1 = 15$	$m_1 = 30$	$m_1 = 5$	$m_1 = 15$	$m_1 = 30$
PFSBS	0.359	0.841	1.022	0.843	0.785	0.595	0.995	1.000	0.994
PFSDDB	0.353	0.854	1.065	0.878	0.937	0.555	0.986	1.036	1.047
FDLB ($k_T = 5$)	0.520	0.591	0.791	1.051	1.006	0.755	1.030	1.068	1.055
FDLB ($k_T = 10$)	**0.266**	0.702	1.002	0.945	0.938	0.520	0.995	1.065	1.047
FDLB ($k_T = 20$)	0.396	0.804	1.075	0.979	0.963	**0.446**	**0.966**	1.051	1.044
FDLB ($k_T = 50$)	0.653	0.926	1.098	0.987	1.010	0.511	0.970	1.045	1.067

Note: The numbers in each cell show the ratio of RMSD obtained with the different bootstrap strategies with respect to the sieve. In bold the lowest ratio for every m.

Table 6: Coverages and widths of confidence intervals: ARFIMA(1,-0.4,0), $\phi = 0.6$

	$T=64$								
	$m=3$			$m=8$			$m=15$		
	$m_1=3$	$m_1=8$	$m_1=15$	$m_1=3$	$m_1=8$	$m_1=15$	$m_1=3$	$m_1=8$	$m_1=15$
Asymp.	**0.947**	**0.947**	**0.947**	0.820	0.820	0.820	0.317	0.317	0.317
	(2.495)	(2.495)	(2.495)	(1.053)	(1.053)	(1.053)	(0.670)	(0.670)	(0.670)
SBS	1.000	1.000	1.000	0.985	0.985	0.985	0.450	0.450	0.450
	(2.689)	(2.689)	(2.689)	(1.239)	(1.239)	(1.239)	(0.774)	(0.774)	(0.774)
PFSBS	0.982	0.998	1.000	0.889	**0.951**	0.932	0.532	0.491	0.409
	(2.469)	(2.597)	(2.721)	(1.238)	(1.239)	(1.235)	(0.793)	(0.777)	(0.768)
PFSDDB	0.981	1.000	1.000	0.817	0.938	0.921	0.583	0.654	0.467
	(2.475)	(2.642)	(2.789)	(1.250)	(1.216)	(1.214)	(0.801)	(0.747)	(0.739)
FDLB ($k_T=3$)	0.939	0.960	0.967	0.775	0.768	0.737	0.393	0.391	0.362
	(1.967)	(2.073)	(2.251)	(1.119)	(1.026)	(1.061)	(0.762)	(0.652)	(0.670)
FDLB ($k_T=5$)	0.982	0.980	0.987	0.811	0.817	0.760	0.520	0.459	0.364
	(2.222)	(2.207)	(2.410)	(1.290)	(1.086)	(1.113)	(0.886)	(0.682)	(0.689)
FDLB ($k_T=10$)	1.000	1.000	0.999	0.820	0.879	0.766	0.624	0.554	0.342
	(2.522)	(2.450)	(2.562)	(1.406)	(1.174)	(1.144)	(1.047)	(0.755)	(0.704)
FDLB ($k_T=25$)	1.000	1.000	1.000	0.812	0.913	0.866	0.677	**0.687**	0.455
	(2.883)	(2.735)	(2.762)	(1.628)	(1.373)	(1.246)	(1.094)	(0.870)	(0.760)
	$T=128$								
	$m=5$			$m=15$			$m=30$		
	$m_1=5$	$m_1=15$	$m_1=30$	$m_1=5$	$m_1=15$	$m_1=30$	$m_1=5$	$m_1=15$	$m_1=30$
Asymp.	0.899	0.899	0.899	0.735	0.735	0.735	0.067	0.067	0.067
	(1.542)	(1.542)	(1.542)	(0.670)	(0.670)	(0.670)	(0.428)	(0.428)	(0.428)
SBS	1.000	1.000	1.000	0.532	0.532	0.532	0.072	0.072	0.072
	(1.862)	(1.862)	(1.862)	(0.742)	(0.742)	(0.742)	(0.467)	(0.467)	(0.467)
PFSBS	0.991	1.000	1.000	0.852	0.808	0.559	0.126	0.101	0.084
	(1.784)	(1.850)	(1.864)	(0.757)	(0.743)	(0.742)	(0.489)	(0.472)	(0.467)
PFSDDB	0.975	1.000	1.000	0.785	0.851	0.547	0.222	0.275	0.053
	(1.764)	(1.858)	(1.876)	(0.765)	(0.738)	(0.735)	(0.493)	(0.461)	(0.455)
FDLB($k_T=5$)	**0.962**	0.976	0.975	0.763	0.735	0.675	0.138	0.114	0.089
	(1.516)	(1.601)	(1.747)	(0.717)	(0.684)	(0.715)	(0.484)	(0.434)	(0.448)
FDLB($k_T=10$)	0.984	0.995	0.995	0.770	0.745	0.561	0.321	0.172	0.069
	(1.708)	(1.702)	(1.852)	(0.831)	(0.706)	(0.730)	(0.557)	(0.449)	(0.456)
FDLB($k_T=20$)	1.000	1.000	1.000	0.730	0.794	0.466	0.554	0.348	0.049
	(1.900)	(1.814)	(1.872)	(0.893)	(0.749)	(0.737)	(0.659)	(0.488)	(0.457)
FDLB($k_T=50$)	0.999	1.000	1.000	0.802	**0.871**	0.592	**0.648**	0.608	0.138
	(2.220)	(2.041)	(1.951)	(1.098)	(0.892)	(0.778)	(0.730)	(0.563)	(0.485)

Note: The first number in each cell is the coverage frequency obtained with 1000 replications. The number between round brackets is the average width of the confidence intervals. In bold the closest coverage to 0.95 nominal coverage for every m.

Table 7: Coverages and widths of confidence intervals: ARFIMA(1,0.4,0), $\phi = 0.6$

	$T = 64$								
	$m = 3$			$m = 8$			$m = 15$		
	$m_1 = 3$	$m_1 = 8$	$m_1 = 15$	$m_1 = 3$	$m_1 = 8$	$m_1 = 15$	$m_1 = 3$	$m_1 = 8$	$m_1 = 15$
Asymp.	0.880	0.880	0.880	0.830	0.830	0.830	0.376	0.376	0.376
	(2.495)	(2.495)	(2.495)	(1.053)	(1.053)	(1.053)	(0.670)	(0.670)	(0.670)
SBS	0.994	0.994	0.994	0.984	0.984	0.984	0.685	0.685	0.685
	(2.929)	(2.929)	(2.929)	(1.260)	(1.260)	(1.260)	(0.796)	(0.796)	(0.796)
PFSBS	0.991	0.997	0.999	0.887	0.933	0.927	0.562	0.512	0.418
	(2.672)	(2.861)	(2.827)	(1.171)	(1.207)	(1.195)	(0.757)	(0.760)	(0.742)
PFSDDB	0.998	1.000	0.999	0.789	0.921	0.915	0.558	0.687	0.529
	(2.715)	(2.882)	(2.809)	(1.205)	(1.192)	(1.178)	(0.768)	(0.733)	(0.718)
FDLB ($k_T = 3$)	0.933	**0.960**	0.971	0.786	0.775	0.776	0.441	0.434	0.401
	(2.041)	(2.061)	(2.162)	(1.163)	(1.028)	(1.061)	(0.771)	(0.643)	(0.662)
FDLB ($k_T = 5$)	0.979	0.979	0.988	0.835	0.826	0.797	0.578	0.492	0.420
	(2.360)	(2.217)	(2.307)	(1.361)	(1.082)	(1.109)	(0.914)	(0.673)	(0.682)
FDLB ($k_T = 10$)	0.999	0.999	0.997	0.832	0.892	0.827	0.661	0.600	0.418
	(2.713)	(2.568)	(2.512)	(1.516)	(1.172)	(1.135)	(1.131)	(0.743)	(0.695)
FDLB ($k_T = 25$)	1.000	1.000	1.000	0.851	**0.934**	0.913	0.676	**0.697**	0.543
	(2.979)	(2.887)	(2.779)	(1.790)	(1.354)	(1.216)	(1.230)	(0.837)	(0.736)
	$T = 128$								
	$m = 5$			$m = 15$			$m = 30$		
	$m_1 = 5$	$m_1 = 15$	$m_1 = 30$	$m_1 = 5$	$m_1 = 15$	$m_1 = 30$	$m_1 = 5$	$m_1 = 15$	$m_1 = 30$
Asymp.	0.891	0.891	0.891	0.752	0.752	0.752	0.092	0.092	0.092
	(1.542)	(1.542)	(1.542)	(0.670)	(0.670)	(0.670)	(0.428)	(0.428)	(0.428)
SBS	0.995	0.995	0.995	**0.944**	**0.944**	**0.944**	0.177	0.177	0.177
	(1.880)	(1.880)	(1.880)	(0.765)	(0.765)	(0.765)	(0.494)	(0.494)	(0.494)
PFSBS	0.972	0.998	1.000	0.831	0.813	0.584	0.174	0.123	0.082
	(1.822)	(1.855)	(1.829)	(0.748)	(0.739)	(0.733)	(0.488)	(0.471)	(0.462)
PFSDDB	0.970	0.996	0.999	0.768	0.841	0.575	0.290	0.336	0.093
	(1.831)	(1.861)	(1.835)	(0.758)	(0.736)	(0.726)	(0.493)	(0.464)	(0.453)
FDLB($k_T = 5$)	**0.956**	0.980	0.982	0.789	0.751	0.699	0.186	0.166	0.142
	(1.614)	(1.655)	(1.762)	(0.723)	(0.684)	(0.712)	(0.484)	(0.432)	(0.445)
FDLB($k_T = 10$)	0.986	0.993	0.994	0.798	0.779	0.631	0.364	0.225	0.119
	(1.865)	(1.762)	(1.851)	(0.853)	(0.704)	(0.725)	(0.560)	(0.444)	(0.450)
FDLB($k_T = 20$)	0.998	1.000	1.000	0.734	0.831	0.544	0.591	0.384	0.094
	(2.124)	(1.879)	(1.871)	(0.958)	(0.746)	(0.729)	(0.676)	(0.477)	(0.451)
FDLB($k_T = 50$)	1.000	1.000	1.000	0.811	0.881	0.667	**0.650**	0.619	0.190
	(2.424)	(2.079)	(1.912)	(1.196)	(0.848)	(0.755)	(0.785)	(0.531)	(0.465)

Note: The first number in each cell is the coverage frequency obtained with 1000 replications. The number between round brackets is the average width of the confidence intervals. In bold the closest coverage to 0.95 nominal coverage for every m.

Table 8: Coverages and widths of confidence intervals: ARFIMA(1,0.7,0), $\phi = 0.6$.

	$T = 64$								
	$m = 3$			$m = 8$			$m = 15$		
	$m_1 = 3$	$m_1 = 8$	$m_1 = 15$	$m_1 = 3$	$m_1 = 8$	$m_1 = 15$	$m_1 = 3$	$m_1 = 8$	$m_1 = 15$
Asymp.	0.928	0.928	0.928	0.852	0.852	0.852	0.509	0.509	0.509
	(2.495)	(2.495)	(2.495)	(1.053)	(1.053)	(1.053)	(0.670)	(0.670)	(0.670)
SBS	0.987	0.987	0.987	0.987	0.987	0.987	**0.855**	**0.855**	**0.855**
	(2.855)	(2.855)	(2.855)	(1.245)	(1.245)	(1.245)	(0.806)	(0.806)	(0.806)
PFSBS	0.986	0.998	1.000	0.909	0.958	0.984	0.674	0.665	0.607
	(2.566)	(2.635)	(2.549)	(1.151)	(1.152)	(1.132)	(0.757)	(0.730)	(0.710)
PFSDDB	0.996	0.999	0.998	0.810	0.956	0.978	0.625	0.733	0.702
	(2.613)	(2.620)	(2.514)	(1.184)	(1.131)	(1.103)	(0.769)	(0.705)	(0.679)
FDLB($k_T = 3$)	0.937	**0.958**	0.971	0.789	0.782	0.792	0.546	0.518	0.496
	(1.963)	(1.908)	(1.962)	(1.138)	(0.992)	(1.020)	(0.731)	(0.616)	(0.630)
FDLB ($k_T = 5$)	0.993	0.981	0.990	0.848	0.843	0.829	0.652	0.566	0.518
	(2.298)	(2.072)	(2.119)	(1.340)	(1.048)	(1.066)	(0.875)	(0.642)	(0.650)
FDLB ($k_T = 10$)	0.999	0.999	0.999	0.870	0.904	0.875	0.695	0.656	0.557
	(2.622)	(2.372)	(2.301)	(1.501)	(1.117)	(1.084)	(1.110)	(0.698)	(0.658)
FDLB ($k_T = 25$)	1.000	1.000	1.000	0.881	0.930	**0.946**	0.696	0.702	0.608
	(2.796)	(2.610)	(2.474)	(1.718)	(1.223)	(1.110)	(1.202)	(0.758)	(0.675)
	$T = 128$								
	$m = 5$			$m = 15$			$m = 30$		
	$m_1 = 5$	$m_1 = 15$	$m_1 = 30$	$m_1 = 5$	$m_1 = 15$	$m_1 = 30$	$m_1 = 5$	$m_1 = 15$	$m_1 = 30$
Asymp.	0.907	0.907	0.907	0.764	0.764	0.764	0.177	0.177	0.177
	(1.542)	(1.542)	(1.542)	(0.670)	(0.670)	(0.670)	(0.428)	(0.428)	(0.428)
SBS	0.990	0.990	0.990	**0.987**	**0.987**	**0.987**	0.283	0.283	0.283
	(1.850)	(1.850)	(1.850)	(0.764)	(0.764)	(0.764)	(0.505)	(0.505)	(0.505)
PFSBS	0.987	0.998	1.000	0.835	0.813	0.695	0.211	0.147	0.114
	(1.757)	(1.765)	(1.701)	(0.722)	(0.701)	(0.681)	(0.477)	(0.455)	(0.443)
PFSDDB	0.985	0.997	0.999	0.782	0.848	0.712	0.299	0.336	0.156
	(1.763)	(1.758)	(1.696)	(0.738)	(0.696)	(0.669)	(0.491)	(0.443)	(0.421)
FDLB($k_T = 5$)	**0.958**	0.980	0.983	0.782	0.740	0.709	0.248	0.218	0.189
	(1.550)	(1.565)	(1.619)	(0.689)	(0.656)	(0.679)	(0.452)	(0.409)	(0.420)
FDLB($k_T = 10$)	0.989	0.995	0.997	0.810	0.767	0.652	0.414	0.260	0.184
	(1.764)	(1.649)	(1.678)	(0.810)	(0.675)	(0.692)	(0.519)	(0.420)	(0.426)
FDLB($k_T = 20$)	0.999	1.000	0.999	0.740	0.810	0.622	0.613	0.406	0.169
	(2.000)	(1.746)	(1.704)	(0.902)	(0.708)	(0.695)	(0.624)	(0.448)	(0.428)
FDLB($k_T = 50$)	1.000	1.000	1.000	0.801	0.844	0.690	**0.624**	0.579	0.247
	(2.219)	(1.878)	(1.724)	(1.088)	(0.771)	(0.702)	(0.710)	(0.481)	(0.433)

Note: The first number in each cell is the coverage frequency obtained with 1000 replications. The number between round brackets is the average width of the confidence intervals. In bold the closest coverage to 0.95 nominal coverage for every m.