Submitted to the Annals of Statistics
arXiv: arXiv:0000.0000

BOOTSTRAP LONG MEMORY PROCESSES IN THE FREQUENCY DOMAIN

BY JAVIER HIDALGO

Houghton Street
London WC2A 2AE
UK
f.j.hidalgo@lse.ac.uk

The aim of the paper is to describe a bootstrap, contrary to the sieve bootstrap, valid under either long memory (LM) or short memory (SM) dependence. One of the reasons of the failure of the sieve bootstrap in our context is that under LM dependence, the sieve bootstrap may not be able to capture the true covariance structure of the original data. We also describe and examine the validity of the bootstrap scheme for the least squares estimator of the parameter in a regression model and for model specification. The motivation for the latter example comes from the observation that the asymptotic distribution of the test is intractable.

1. INTRODUCTION.

Inference on statistics of interest is often carried out by employing the asymptotic distribution as an approximation to their finite sample one. However such an approximation is not always satisfactory and researchers have then looked for alternative approaches. One of them is resampling methods, introduced by Efron's seminal paper (1979) for independent and identically distributed, *iid*, data. Motivated by its finite sample refinements and statistical properties in different contexts, see for instance Hall (1992), Efron's resampling ideas have been extended to a variety of different *non-iid* situations, including dependent data. The validity of the different bootstrap algorithms depend crucially on the dependence structure of the data and/or on the statistic under consideration. See for instance Bühlmann (2002), Lahiri (2003) or Politis (2003) among others for comprehensive reviews on resampling dependent data.

When resampling dependent data we can differentiate two main methodologies. A first methodology is based on time domain methods, being the two most common ones the Moving Block Bootstrap (MBB), see Künsch (1989), and the $AR-$sieve bootstrap, introduced by Kreiss (1988) and explored by Bühlmann (1997). Among these two approaches, due to its computationally and conceptual simplicity, the AR-sieve bootstrap is very much employed with real data and its validity has been shown for a variety of statistics for time series sequences not necessarily being linear. We refer to Kreiss, Paparoditis and Politis (2011) for a thorough discussion when the $AR-$sieve bootstrap is expected to be valid, who also gave situations for its nonvalidity, such as when the data is noncausal, whereas under long memory (LM) dependence they cast some doubts on its validity. Notice that the latter type of dependence does not satisfy their conditions. In addition to Kreiss et al.'s (2011) comments, we have some further doubts on the validity of the $AR-$sieve bootstrap. The reason being as the bootstrap may not preserve the covariance structure of the data, which is one of the key requirements for the validity of any resampling scheme. Indeed the $AR-$sieve bootstrap appears to match, if anything, a Type II dependence structure whereas Condition 1, see Section 2, implies that u_t may have a Type I dependence. See Marinucci and Robinson's (1999)

MSC 2010 subject classifications: Primary 62G10, 62F40; secondary 62J20, 62G30
Keywords and phrases: long memory, bootstrap methods

Type I and II definitions for fractional Gaussian motions. See also Remark 4 for some extra arguments and/or comments.

A second (general) approach to implement resampling schemes is based on frequency domain methods. The idea or motivation behind this approach comes from the observation that periodogram ordinates at a finite number of frequencies are approximately independent and exponentially distributed, see Brockwell and Davis's (1991) Theorem 10.3.1., so that Efron's ideas may be employed. Early examples are Franke and Härdle's (1992) bootstrap of the spectral density function or Dahlhaus and Janas (1996) for statistics based on functionals of the periodogram. A similar approach might be based on resampling the discrete Fourier transform, *DFT*, as similarly to the periodogram, the *DFT* at two different frequencies can be considered approximately independent. See Hurvich and Zeger (1987) who proposed a nonparametric bootstrap although without any theoretical justification, or Hidalgo (2003) who showed the validity of the resampling for the least squares estimator in a time series regression model. More recently Kirsch and Politis (2011), who also gave a comprehensive review of the literature, proposed and examined a bootstrap scheme with the aim to obtain time series resamples by using the inverse of the *DFT*, which they called *Time Frequency Toggle* (TFT).

However, the previous frequency domain resampling schemes have some limitations. To see this, suppose that $\{u_t\}_{t\in\mathbb{Z}}$ is a sequence of random variables which has an $MA(\infty)$ representation

$$u_t = \sum_{j=0}^{\infty} \vartheta_j \varepsilon_{t-j}; \text{ with } \vartheta_0 = 1,$$

where $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a zero mean *iid* sequence of random variables with $E\left(\varepsilon_t^2\right) = \sigma_\varepsilon^2$. Basically what the aforementioned frequency domain bootstrap schemes do is to "approximate" $\left|\sum_{j=0}^{\infty} \vartheta_j e^{ij\lambda}\right|$, the modulus of the spectral transfer function, so that it suggests that the scheme will be valid only for statistics that only require to mimic the second order dependence structure of the data. So, they might not be a valid resampling scheme for the estimators of the parameters in the transformed regression model

$$\Xi\left(y_t;\delta\right) = \beta' x_t + u_t,$$

such as the well employed Box-Cox transformation $\Xi\left(y_t;\delta\right) =: \frac{y_t^\delta - 1}{\delta}$, as their statistical properties involve features beyond the second moments of the data.

The aim of the paper is thus to describe and examine a bootstrap algorithm which, unlike the methods mentioned previously, approximates directly the transfer function $\sum_{j=0}^{\infty} \vartheta_j e^{ij\lambda}$ as opposed to its modulus, and hence able to match moments greater than or equal to 2 of the data as it is the case with the AR-sieve bootstrap scheme, see Bühlmann (1997). However contrary to the $AR-$sieve bootstrap, we want to allow for sequences that might exhibit LM dependence, as well as short memory (SM) dependence. Also, as our resampling scheme obtains time series resamples, it has some similarities with the TFT scheme, and thus it appears that the conditions in Kirsch and Politis (2011) might be sufficient but not necessary. Observe that Kirsch and Politis's (2011) Assumptions $\mathcal{P}.\mathbf{1}$ to $\mathcal{P}.\mathbf{3}$ are not satisfied for LM dependent sequences.

The bootstrap described in Section 2 is based on the "discrete" Cramér representation of $\mathcal{U}_n = \{u_t\}_{t=1}^n$ and Bartlett's approximation of the *discrete Fourier transform* of $\mathcal{U}_n$ by that of its innovation sequence, say $\{\varepsilon_t\}_{t=1}^n$. In addition, our resampling scheme, contrary to the *MBB*, is not a subset of the original data and similar to the TFT scheme, we obtain time series resamples. Finally, similar to the AR-sieve bootstrap scheme, the bootstrap data sequence is covariance stationary.

The remainder of the paper is as follows. Section 2 states the regularity conditions and it describes the resampling scheme. Section 3 describes two main situations where the proposed bootstrap scheme is valid. More specifically, for the least squares estimator in a time series regression model and for model specification in regression models when both regressors and error term may exhibit LM dependence. Section 4 reports the results of a Monte Carlo study of the finite sample performance of the bootstrap, and in the case of SM sequences how it compares with the AR-sieve bootstrap, whereas Section 5 concludes. The proofs of all our main results in Section 2 are collected in Section 6, whereas Section 7 states a series of Lemmas for easy reference. The supplementary material provides the proofs of the main results as well as the lemmas.

2. REGULARITY CONDITIONS AND DESCRIPTION OF THE BOOTSTRAP.

We give some notation first. For any $d \in (-1/2, 1/2)$, we denote

$$(1) \qquad (1-L)^d = \sum_{k=0}^{\infty} \pi_k(d) L^k; \quad \pi_k(-d) = \frac{\Gamma(k+d)}{\Gamma(d)\Gamma(k+1)}, \quad k \in \mathbb{N}$$

$$\left(1 - 2\cos\omega L + L^2\right)^d = \sum_{k=0}^{\infty} \tau_k(\cos\omega; d) L^k,$$

where $\Gamma(\cdot)$ denotes the gamma function such that $\Gamma(c) = \infty$ for $c = 0$ with $\Gamma(0)/\Gamma(0) = 1$ and the coefficients $\tau_k(\cos\omega; d)$ follow the second order homogeneous difference equation

$$\tau_k(z; d) = 2z\left(\frac{k-d-1}{k}\right)\tau_{k-1}(z; d) - \left(\frac{k-2d-2}{k}\right)\tau_{k-2}(z; d),$$

see Section 8.93 in Gradshteyn and Ryzhik (2000).

We now introduce the following regularity condition.

CONDITION 1. *$\{u_t\}_{t\in\mathbb{Z}}$ is a sequence of random variables such that*

$$u_t = \sum_{j=0}^{\infty} \vartheta_j \varepsilon_{t-j}; \quad \sum_{j=0}^{\infty} \vartheta_j^2 < \infty, \quad \vartheta_0 = 1,$$

where $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a zero mean iid *sequence of random variables with $E\left(\varepsilon_t^2\right) = \sigma_\varepsilon^2$ and a spectral density function, $f_u(\lambda)$, bounded away from zero. Also*

$$(2) \qquad \vartheta_j = \sum_{k=0}^{j} \xi_k(-d_1; -d_2) b_{k-j} = \sum_{k=0}^{j} \xi_{k-j}(-d_1; -d_2) b_k; \qquad d_1, d_2 \in [0, 1/2),$$

where $\xi_k(d_1; d_2) = \sum_{\ell=0}^{k} \tau_\ell(\cos\omega; d_2)\pi_{k-\ell}(d_1) =: \sum_{\ell=0}^{k} \tau_{k-\ell}(\cos\omega; d_2)\pi_\ell(d_1)$ and $\sum_{k=0}^{\infty} k^2 |b_k| < \infty$.

Condition 1 implies that the process $\{u_t\}_{t\in\mathbb{Z}}$ belongs to the class of linear processes and it implies that for instance

$$u_t = \begin{cases} \sum_{j=0}^{\infty} \xi_j(-d_1; -d_2)\varepsilon'_{t-j}; \quad \varepsilon'_t = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j} & \text{if } d_1, d_2 > 0 \\ \sum_{j=0}^{\infty} b_j \varepsilon_{t-j} & \text{if } d_1 = d_2 = 0 \end{cases}$$

and denoting

$$(3) \qquad g_1(\lambda, d) =: \left(1 - e^{-i\lambda}\right)^d = \sum_{j=0}^{\infty} \pi_j(d) e^{-ij\lambda}$$

$$g_2(\lambda, d) =: \left(1 - 2\cos\omega e^{-i\lambda} + e^{-i2\lambda}\right)^d = \sum_{j=0}^{\infty} \tau_j(\cos\omega; d) e^{-ij\lambda}$$

$$B(\lambda) = \sum_{j=0}^{\infty} b_j e^{-ij\lambda}; \qquad \lambda \in (-\pi, \pi],$$

the decomposition given in (2), together with (3), implies that $f_u(\lambda)$ can be factorized as

$$f_u(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi} |g(\lambda, -d_1, -d_2)|^2 |B(\lambda)|^2; \quad \lambda \in (-\pi, \pi], \tag{4}$$

$$g(\lambda, d_1, d_2) = g_1(\lambda, d_1) g_2(\lambda, d_2) =: \sum_{j=0}^{\infty} \xi_j(d_1, d_2) e^{-ij\lambda}.$$

Recall that our condition $\sum_{k=0}^{\infty} k^2 |b_k| < \infty$ implies that $|B(\lambda)|^2$ is twice continuously differentiable for all $\lambda \in [0, \pi]$. Observe that the condition $d_1, d_2 \geq 0$ implies that $f_u(\lambda) > 0$ for any $\lambda \in [0, \pi]$. Also the sequence $\{u_t\}_{t\in\mathbb{Z}}$ admits the AR representation

$$u_t = \sum_{j=1}^{\infty} \phi_j u_{t-j} + \varepsilon_t; \qquad \phi_j =: \sum_{k=0}^{j} \xi_k(d_1; d_2) a_{k-j} =: \sum_{k=0}^{j} \xi_{k-j}(d_1; d_2) a_k, \tag{5}$$

where $\sum_{k=0}^{\infty} k^2 |a_k| < \infty$ and $B^{-1}(\lambda) =: A(\lambda) = \sum_{j=0}^{\infty} a_j e^{-ij\lambda}$. So, we can then write also the spectral density function as

$$f_u(\lambda) =: \frac{\sigma_\varepsilon^2}{2\pi} |g(\lambda; d_1, d_2)|^{-2} |A(\lambda)|^{-2}.$$

One model satisfying (2) is the multiplicative $GARMA(p, d_1, d_2, q)$ process

$$(1 - L)^{d_1} \left(1 - 2\cos\omega L + L^2\right)^{d_2} \Phi_p(L) u_t = \Theta_q(L) \varepsilon_t,$$

where $\Phi_p(L)$ and $\Theta_q(L)$ are the autoregressive and moving average polynomials with no common roots and outside the unit circle. The latter implies that $\Phi_p^{-1}(L) \Theta_q(L) = \sum_{j=0}^{\infty} b_j L^j$ with $b_j = O(j^{-\upsilon})$ for any $\upsilon > 0$.

It is worth mentioning that the results hold true under the slightly weaker condition and where for the sake of the argument, we shall consider the case $d_2 = 0$.

CONDITION 2. *$\{u_t\}_{t\in\mathbb{Z}}$ is a sequence of random variables such that*

$$u_t = \sum_{j=0}^{\infty} \vartheta_j \varepsilon_{t-j}; \quad \sum_{j=0}^{\infty} \vartheta_j^2 < \infty, \quad \vartheta_0 = 1,$$

where $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a zero mean sequence of iid *random variables and $E\left(\varepsilon_t^2\right) = \sigma_\varepsilon^2$. Also*

$$\vartheta_j = \sum_{k=0}^{j} \overline{\vartheta}_k b_{k-j},$$

where $\overline{\vartheta}_k = \ell(k) k^{d_1 - 1}$; $|\ell(k) - \ell(k+1)| < \ell'(k) k^{-1}$ with $\ell'(k) > 0$, $d_1 \in \left[0, \frac{1}{2}\right)$ and $\sum_{k=0}^{\infty} k^2 |b_k| < \infty$.

Condition 2 is similar to that in Marinucci and Robinson (2000). An example of $\ell(k)$ is any slowly varying function. Moreover, because Condition 2 implies that the sequence $\left\{\overline{\vartheta}_k\right\}_{k=0}^{\infty}$ is of bounded variation, that is $\sum_{k=0}^{\infty} \left|\overline{\vartheta}_k - \overline{\vartheta}_{k-1}\right| < \infty$, and because $\left\{\overline{\vartheta}_k\right\}_{k=0}^{\infty}$

is a quasi-monotonic sequence, see Yong $(1974, \text{p.}2)$ for a definition, we then have that $\left|\sum_{k=0}^{\infty}\overline{\vartheta}_k e^{-ik\lambda}\right|^2 \sim D\lambda^{-2d_1}$ as $\lambda \to 0+$, $0 < D < \infty$, and continuous differentiable outside any open set around the zero frequency, see Yong's (1974) Theorems III-11 and 12. However, we have preferred to keep Condition 1 as it eases the notation and it does not unnecessarily complicate and lengthen the arguments of our results.

Before we present our bootstrap scheme, it is also worth given the trivial observation, which plays an important role in the bootstrap scheme below, that

(6)
$$g^{-1}(\lambda; d_1, d_2) =: \left(1 - e^{-i\lambda}\right)^{-d_1}\left(1 - 2\cos\omega e^{-i\lambda} + e^{-i2\lambda}\right)^{-d_2} = \sum_{j=0}^{\infty}\xi_j(-d_1, -d_2)e^{-ij\lambda} =: g(\lambda; -d_1, -d_2).$$

In what follows the upperscript "$*$" denotes the bootstrap analogue. That is, P^* and $E^*(z)$ denote respectively the probability and the expectation conditional on $\{u_t\}_{t=1}^n$. We shall now describe and examine a valid bootstrap scheme when the data may exhibit either LM or SM dependence. For that purpose, for a generic sequence $\{z_t\}_{t=1}^n$, we write the *DFT* and *periodogram* respectively as

(7)
$$w_z(\lambda) = \frac{1}{n^{1/2}}\sum_{t=1}^{n} z_t e^{-it\lambda}; \quad I_{zz}(\lambda) = |w_z(\lambda)|^2.$$

Also, we denote the Fourier frequencies as $\lambda_j = 2\pi j/n$, for integer $j \geq 0$.

The idea behind the resampling scheme is based on two well known results. First the identity

(8)
$$u_t = \frac{1}{n^{1/2}}\sum_{j=1}^{n} e^{it\lambda_j} w_u(\lambda_j), \quad t = 1, ..., n,$$

and secondly the Bartlett's approximation of the *DFT* of $\{u_t\}_{t=1}^n$ by that of $\{\varepsilon_t\}_{t=1}^n$. That is,

(9)
$$w_u(\lambda_j) \approx g^{-1}(\lambda_j; d_1, d_2) B(\lambda_j) w_\varepsilon(\lambda_j),$$

or using (6)

(10)
$$w_u(\lambda_j) \approx g(\lambda_j; -d_1, -d_2) B(\lambda_j) w_\varepsilon(\lambda_j),$$

where "$\approx$" should be read as "approximately". Expressions (8) and (10) suggest that we might approximate u_t by

(11)
$$u_t \approx \frac{1}{n^{1/2}}\sum_{j=1}^{n} e^{it\lambda_j} g(\lambda_j; -d_1, -d_2) B(\lambda_j) w_\varepsilon(\lambda_j).$$

However the existence of a singularity for $g(\lambda; -d_1, -d_2) =: g^{-1}(\lambda; d_1, d_2)$ at $\lambda = 0$ when $d_1 > 0$ and at $\lambda = \omega$ when $d_2 > 0$, for instance, suggests that the approximation given in (10), or (9), is not adequate for frequencies near zero or ω, see for instance Robinson $(1995a)$. In addition, see Remark 4 below for some details, the implicit circularity induced on the right side of (11) casts some additional doubts on the validity of the approximation given in (11). So, the idea of the bootstrap scheme is based on a modification of (11) given by

(12)
$$u_t \approx \frac{1}{n^{1/2}}\sum_{j=1}^{n} e^{it\lambda_j} g(\lambda_j; -d_1, -d_2; L) B(\lambda_j) w_\varepsilon(\lambda_j),$$

where, using (4), for any $d_1, d_2 \in (-1/2, 1/2)$,

(13) $$g(\lambda; d_1, d_2; L) = \sum_{j=0}^{L} \xi_j (d_1, d_2) e^{-ij\lambda}.$$

Thus, for instance we can view the modification given in (12) as a trimming version of (11). It is worth observing that (6) and (13) yield that

$$g(\lambda; d_1, d_2; \infty) = g^{-1}(\lambda; -d_1, -d_2; \infty) =: \left(1 - e^{-i\lambda}\right)^{d_1} \left(1 - 2\cos\omega e^{-i\lambda} + e^{-i2\lambda}\right)^{d_2},$$

so that we can then approximate $g(\lambda; -d_1, -d_2; \infty)$ by either $g^{-1}(\lambda; d_1, d_2; L)$ or $g(\lambda; -d_1, -d_2; L)$. Although there is no doubts that the results hold true using either approximation, we shall employ $g(\lambda; -d_1, -d_2; L)$ or $g(\lambda; d_1, d_2; L)$ if our purpose is respectively to approximate u_t, i.e. (11), or ε_t, i.e.

$$\varepsilon_t \approx \frac{1}{n^{1/2}} \sum_{j=1}^{n} e^{it\lambda_j} g(\lambda_j; d_1, d_2; L) A(\lambda_j) w_u(\lambda_j).$$

The motivation is because it simplifies significantly the computation of the bootstrap, see *STEPS 1&3* below, as well as some of the technical aspects of the proofs.

Now (12) suggests that if d_1, d_2 and $B(\lambda_j)$ were replaced respectively by consistent estimators $\widehat{d}_1, \widehat{d}_2$ and $\widehat{B}(\lambda_j)$, the problem to obtain a valid bootstrap sample $\{u_t^*\}_{t=1}^n$ becomes a problem of designing a valid bootstrap algorithm for $w_\varepsilon(\lambda_j)$, $j = 1, ..., \tilde{n} = [n/2]$. Regarding the estimator of d_1, we shall employ Robinson's $(1995b)$ Local Pseudo-Gaussian estimator. That is,

(14) $$\widehat{d}_1 = \arg \min_{d \in [0,\Delta]} \left\{ \log \left(\frac{1}{m} \sum_{j=1}^{m} \lambda_j^{2d} I_{uu}(\lambda_j) \right) - \frac{2d}{m} \sum_{j=1}^{m} \log \lambda_j \right\},$$

where $0 < \Delta < 1/2$, $m = m(n)$ increasing with the sample size n. On the other hand to estimate d_2, we employ an obvious extension of (14), see Arteche (2000) or Hidalgo (2005) for an alternative method, given by

(15)
$$\widehat{d}_2 = \arg \min_{d \in [0,\Delta]} \left\{ \log \left(\frac{1}{2m} \sum_{j=-m;\neq 0}^{m} |\lambda_j - \omega|^{2d} I_{uu}(\lambda_j) \right) - \frac{2d}{2m} \sum_{j=-m;\neq 0}^{m} \log |\lambda_j - \omega| \right\}.$$

Also we estimate $h(\lambda) =: g(\lambda; d_1, d_2) f_u(\lambda)$ by

(16) $$\widehat{h}(\lambda) = \frac{1}{2m+1} \sum_{k=-m}^{m} \left| g\left(\lambda + \lambda_k; \widehat{d}_1, \widehat{d}_2\right) \right|^2 \frac{I_{uu}(\lambda + \lambda_k)}{2\pi},$$

which can be regarded as an estimator of $\sigma_\varepsilon^2 |B(\lambda)|^2 / 2\pi = \sigma_\varepsilon^2 |A(\lambda)|^{-2} / 2\pi$. As a by-product

$$\widehat{f}_u(\lambda) = \left| g\left(\lambda; \widehat{d}_1, \widehat{d}_2\right) \right|^{-2} \widehat{h}(\lambda)$$

becomes an estimator of $f_u(\lambda)$, which can be viewed as a prewhitening and then recolouring type of estimator for the spectral density function, see Press and Tukey (1956).

Before we present the bootstrap scheme, denoting $M = [n/4m]$ we introduce

$$\widehat{C}(\lambda) = \exp\left\{\sum_{\ell=1}^{M} \widehat{c}_\ell e^{-i\ell\lambda}\right\}, \tag{17}$$

$$\widehat{c}_\ell = \frac{1}{M}\sum_{j=1}^{M} \log \widehat{h}(\lambda_{2mj}) \cos \ell\lambda_{2mj}, \quad \ell = 0, ..., M,$$

with $\widehat{h}(\lambda_j)$ given in (16). Recall that $\widehat{\sigma}_\varepsilon^2 =: 2\pi \exp(\widehat{c}_0)$ is an estimator of σ_ε^2. Finally in what follows, for any integer $k \geq 0$,

$$\begin{aligned}
\pi_k &=: \pi_k(d_1);\ \widehat{\pi}_k =: \pi_k\left(\widehat{d}_1\right); \\
\overline{\pi}_k &=: \pi_k(-d_1);\ \check{\pi}_k =: \pi_k\left(-\widehat{d}_1\right) \\
\tau_k &=: \tau_k(\cos\omega; d_2);\ \widehat{\tau}_k =: \tau_k\left(\cos\omega; \widehat{d}_2\right); \\
\overline{\tau}_k &=: \tau_k(\cos\omega; -d_2);\ \check{\tau}_k =: \tau_k\left(\cos\omega; -\widehat{d}_2\right) \\
\xi_k &=: \xi_k(d_1, d_2);\ \widehat{\xi}_k =: \xi_k\left(\widehat{d}_1, \widehat{d}_2\right) \\
\overline{\xi}_k &=: \xi_k(-d_1, -d_2);\ \check{\xi}_k =: \xi_k\left(-\widehat{d}_1, -\widehat{d}_2\right).
\end{aligned} \tag{18}$$

Our bootstrap scheme is given in the following *3 STEPS*.

STEP 1 We compute the innovations as

$$\widehat{\varepsilon}_t = \frac{1}{n^{1/2}}\sum_{j=1}^{n} e^{it\lambda_j}\widehat{\Phi}(\lambda_j) w_u(\lambda_j), \quad t = 1, ..., n, \tag{19}$$

$$\widehat{\Phi}(\lambda) = g\left(\lambda; \widehat{d}_1, \widehat{d}_2; n - M\right)\widehat{A}(\lambda)$$

with $g(\lambda; d_1, d_2; L)$ given in (13) and $\widehat{A}(\lambda) = \sum_{k=0}^{M} \widehat{a}_k e^{-ik\lambda}$, where $\widehat{a}_0 = 1$ and

$$\widehat{a}_k = \frac{1}{2M+1}\sum_{j=-M}^{M} \widehat{C}^{-1}(\lambda_{2mj}) e^{-ik\lambda_{2mj}}, \quad k = 1, ..., M.$$

REMARK 1. (i) *Standard algebra yields that* $\widehat{\Phi}(\lambda) = \sum_{\ell=0}^{n} \widehat{\phi}_\ell e^{-i\ell\lambda_j}$, *where* $\widehat{\phi}_0 = 1$ *and*

$$\widehat{\phi}_\ell = \begin{cases} \sum_{k=0}^{\ell\wedge M} \widehat{\xi}_{\ell-k}\widehat{a}_k, & 1 \leq \ell \leq n - M \\ \sum_{k=\ell-(n-M)}^{M} \widehat{\xi}_{\ell-k}\widehat{a}_k, & n - M < \ell \leq n. \end{cases} \tag{20}$$

(ii) *Because for any sequence* $\{\nu_\ell\}_{\ell=0}^{n}$,

$$\frac{1}{n^{1/2}}\sum_{j=1}^{n} e^{it\lambda_j}\left(\sum_{\ell=0}^{n} \nu_\ell e^{-i\ell\lambda_j}\right) w_z(\lambda_j) =: \sum_{\ell=0}^{t-1} \nu_\ell z_{t-\ell} + \sum_{\ell=t}^{n} \nu_\ell z_{n-(\ell-t)}, \tag{21}$$

we have that $\widehat{\varepsilon}_t$ *in* (19) *can be written as*

$$\widehat{\varepsilon}_t = \sum_{\ell=0}^{t-1} \widehat{\phi}_\ell u_{t-\ell} + \sum_{\ell=t}^{n} \widehat{\phi}_\ell u_{n+t-\ell}. \tag{22}$$

STEP 2 Draw a random sample $\varepsilon^* = \{\varepsilon^*_t\}_{t=1}^{3n}$ from the empirical distribution of $\{\check{\varepsilon}_t\}_{t=1}^n =: \left\{\widehat{\varepsilon}_t - n^{-1}\sum_{t=1}^n \widehat{\varepsilon}_t\right\}_{t=1}^n$, and compute the *DFT* of ε^*, that is

$$w_{\varepsilon^*}\left(\widetilde{\lambda}_j\right) = \frac{1}{(3n)^{1/2}}\sum_{t=1}^{3n} \varepsilon^*_t e^{-it\widetilde{\lambda}_j}; \quad \widetilde{\lambda}_j = 2\pi j/3n, \quad j = 1, \ldots, 3n.$$

REMARK 2. (i) *We shall mention that we could have computed* $w_{\varepsilon^*}\left(\widetilde{\lambda}_j\right)$ *as*

$$w_{\varepsilon^*}\left(\widetilde{\lambda}_j\right) = \frac{1}{(3n)^{1/2}}\sum_{t=1}^{3n} u^*_t e^{-it\widetilde{\lambda}_j},$$

where $\{u^*_t\}_{t=1}^{3n}$ *is a random sample from the empirical distribution of* $\{\widetilde{u}_t\} = \left\{u_t - n^{-1}\sum_{t=1}^n u_t\right\}_{t=1}^n$. *However, we prefer our* STEP 2 *since as Corollary 2 below shows, we would be able to match the moments of* $\{u_t\}_{t\in\mathbb{Z}}$, *whereas the approach given in the last displayed expression would only match the first two moments of* u_t.

(ii) *Our results hold true if instead of a sample of size* $3n$ *we would have chosen* $2n$. *However, we have decided to choose* $3n$ *instead of say* n *or* $2n$ *for notational simplicity as it becomes clear from expressions* (24) *and* (25) *below.*

(iii) *In* STEP 1 *we could have computed* $\widehat{\Phi}(\lambda) = g\left(\lambda; \widehat{d}_1, \widehat{d}_2; n\right)\widehat{A}(\lambda)$ *instead of* $\widehat{\Phi}(\lambda)$ *as given in* (19). *However, the former introduces some additional notational complication due to end effects, so we have decided for clarity to keep* STEP 1 *as it stands.*

STEP 3 Denoting $\widehat{\Psi}(\lambda) = g\left(\lambda; -\widehat{d}_1, -\widehat{d}_2; n\right)\widehat{B}(\lambda)$ with $\widehat{B}(\lambda) = \sum_{k=0}^M \widehat{b}_k e^{-ik\lambda}$, where

$$\widehat{b}_k = \frac{1}{2M+1}\sum_{j=-M}^{M} \widehat{C}(\lambda_{2mj})\, e^{-ik\lambda_{2mj}}, \qquad k = 1, \ldots M,$$

and $\widehat{b}_0 = 1$, we compute

$$\ddot{u}^*_t = \frac{1}{(3n)^{1/2}}\sum_{j=1}^{3n} e^{it\widetilde{\lambda}_j}\,\widehat{\Psi}\left(\widetilde{\lambda}_j\right) w_{\varepsilon^*}\left(\widetilde{\lambda}_j\right), \quad t = 1, \ldots, 3n \tag{23}$$

and denote our bootstrap sequence as $\{u^*_t\}_{t=1}^n =: \left\{\ddot{u}^*_{t+2n}\right\}_{t=1}^n$.

REMARK 3. (i) $\widehat{\Psi}(\lambda)$ *is an estimator of* $\Psi(\lambda) = g(\lambda; -d_1, -d_2)\,B(\lambda)$ *as is* $\widehat{\Phi}(\lambda)$ *an estimator of* $\Phi(\lambda) = g(\lambda; d_1, d_2)\,A(\lambda)$ *for* $\lambda \in (0,\omega)\cup(\omega,\pi]$.

(ii) *Because* $\sum_{j=1}^{3n} e^{i\ell\widetilde{\lambda}_j} = 0$ *if* $\ell \neq 0, 3n, \ldots$, *we can write* $\ddot{u}^*_t$ *in* (23) *as*

$$\ddot{u}^*_t = \begin{cases} \sum_{\ell=0}^{n+M}\widehat{\vartheta}_\ell \varepsilon^*_{t-\ell} & \text{if} \quad n+M \le t \\ \sum_{\ell=0}^{t-1}\widehat{\vartheta}_\ell \varepsilon^*_{t-\ell} + \sum_{\ell=t}^{n+M}\widehat{\vartheta}_\ell \varepsilon^*_{3n+t-\ell} & \text{if} \quad t < n+M, \end{cases} \tag{24}$$

with $\widehat{\vartheta}_0 = 1$ *and*

$$\widehat{\vartheta}_\ell = \begin{cases} \sum_{p=0}^{\ell\wedge M}\widehat{b}_p \check{\xi}_{\ell-p}, & 1 \le \ell \le n \\ \sum_{p=\ell-n}^{M}\widehat{b}_p \check{\xi}_{\ell-p}, & n < \ell \le n+M. \end{cases} \tag{25}$$

So, the sequence $\{u^*_t\}_{t=1}^n$ *behaves as a* $MA(n+M)$ *process with weights* $\left\{\widehat{\vartheta}_\ell\right\}_{\ell=0}^{n+M}$.

REMARK 4. (i) *Regarding our definition in* (13), *we might be tempted to use* (13) *with* $L=\infty$, *there. However due to the circularity of* $e^{-ij\lambda}$, *it is not feasible. Take* $B(\lambda)=1$ *and* $d_2=0$ *for simplicity. Then, because for any integer* $q\geq 0$, $e^{-i\ell\widetilde{\lambda}_j}=e^{-i(\ell+q3n)\widetilde{\lambda}_j}$, *it implies that if we employed* $g_1(\lambda,d_1;\infty)$, *we would then have that*

$$u_t^*=\sum_{\ell=0}^{3n}\left(\sum_{q=1}^{\infty}\breve{\pi}_{\ell+q3n}\right)\sum_{s=1}^{3n}\varepsilon_s^*\mathcal{I}\left(s=t-\ell \bmod 3n\right),\quad t=1,...,3n.$$

But when $d_1>0$ *the sequence* $\{\overline{\pi}_\ell\}_{\ell\in\mathbb{Z}}$ *is not summable which implies that* $\sum_{q=1}^{\infty}\breve{\pi}_{\ell+q3n}$ *will not be bounded in probability and hence the procedure invalid.*

(ii) *One reason why the* AR*-sieve bootstrap might not be valid under* LM *dependence is due to the fact that it will not catch the singularity of* $f_u(\lambda)$ *at zero and/or* ω *frequencies. Recall that* $\left|1-\sum_{p=1}^{\mathbf{p}_n}\widehat{\alpha}_p e^{-ip\lambda}\right|^{-2}$ *is an estimator of the spectral density function, see Berk* (1974) *or Bühlmann* (1997), *where* $\{\widehat{\alpha}_p\}_{p=1}^{\mathbf{p}_n}$ *are the estimators of the parameters in the* $AR(\mathbf{p}_n)$ *sieve approximation.*

We need to impose some restrictions on the bandwidth m employed to estimate d_1 and d_2 in (14) and (15) respectively.

CONDITION 3. *As* $n\to\infty$,

$$\frac{m^4}{n^3}+\frac{n^2}{m^3}\log n\to 0.$$

Condition 3 gives upper and lower bounds on the rate of increase to infinity of the smoothing parameter m. For example, $m=n^{\psi}$ would satisfy Condition 3 for any $\psi\in\left(\frac{2}{3},\frac{3}{4}\right)$.

Our first two corollaries shows that our bootstrap scheme is able to match the moments of the innovation sequence $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ and those of $\{u_t\}_{t\in\mathbb{Z}}$.

PROPOSITION 1. *Under Conditions* 1 *and* 3, *for any* $p\geq 1$ *such that* $E|\varepsilon_t|^p<\infty$, *we have that*

$$\frac{1}{n}\sum_{t=1}^{n}|\breve{\varepsilon}_t-\varepsilon_t|^p=o_p(1). \tag{26}$$

As a consequence of Proposition 1, we have the following corollary.

COROLLARY 1. *Assume Conditions* 1 *and* 3. *Then, for any* $p\geq 1$ *such that* $E|\varepsilon_t|^p<\infty$, *we have that*

$$E^*\varepsilon_t^{*p}=:\frac{1}{n}\sum_{t=1}^{n}\breve{\varepsilon}_t^p\overset{P}{\to}E\varepsilon_t^p.$$

PROOF. By standard equalities, we have that

$$\frac{1}{n}\sum_{t=1}^{n}\breve{\varepsilon}_t^p=\frac{1}{n}\sum_{t=1}^{n}(\breve{\varepsilon}_t-\varepsilon_t)^p+\sum_{k=1}^{p-1}\frac{1}{n}\sum_{t=1}^{n}\binom{p}{k}(\breve{\varepsilon}_t-\varepsilon_t)^k\varepsilon_t^{p-k}+\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t^p.$$

The first term is $o_p(1)$ by Proposition 1. The third term converges in probability to $E\varepsilon_t^p$ by Condition 1. From here the conclusion follows by Hölder's inequality. ∎

PROPOSITION 2. *Under Conditions* 1 *and* 3*, for any* $p \geq 1$ *such that* $E|\varepsilon_t|^p < \infty$*, we have that*

$$E^* |u_t^* - \widetilde{u}_t^*|^p = o_p(1), \tag{27}$$

where $\widetilde{u}_t^* = \sum_{\ell=0}^{n+M} \vartheta_\ell \varepsilon_{t-\ell}^*$.

Before we state our next corollary, it is worth recalling that for any sequence $\{\zeta_j\}_{j\geq 1}$ and a martingale difference sequence $\{\eta_j\}_{j\in\mathbb{Z}}$ with finite p moments, we have that

$$E\left|\sum_{j=a}^{b} \zeta_j \eta_j\right|^p \leq K\left|\sum_{j=a}^{b} \zeta_j^2 E|\eta_j|^2\right|^{p/2} \leq KE\left|\sum_{j=a}^{b} \zeta_j^2\right|^{p/2-1} \sum_{j=a}^{b} \zeta_j^2 E|\eta_j|^p \tag{28}$$

by Burkholder and then Hölder's inequalities.

A consequence of Proposition 2 is the following corollary.

COROLLARY 2. *Under Conditions* 1 *and* 3 *for any* $p \geq 1$ *such that* $E|\varepsilon_t|^p < \infty$*, we have that*

$$E^* u_t^{*p} - E u_t^p = o_p(1). \tag{29}$$

PROOF. Because $u_t = \widetilde{u}_t + \sum_{\ell=n+M+1}^{\infty} \vartheta_\ell \varepsilon_{t-\ell}$, where $\widetilde{u}_t = \sum_{\ell=0}^{n+M} \vartheta_\ell \varepsilon_{t-\ell}$, and using (28) and that $\{\vartheta_\ell^2\}_{\ell\geq 1}$ is summable, $E\left|\sum_{\ell=n+M+1}^{\infty} \vartheta_\ell \varepsilon_{t-\ell}\right|^p = o(1)$, it suffices to show (29) but with u_t replaced by $\widetilde{u}_t$. Now, by standard equalities, we have that

$$u_t^{*p} - \widetilde{u}_t^{*p} = (u_t^* - \widetilde{u}_t^*)^p + \sum_{k=1}^{p-1} \binom{p}{k} (u_t^* - \widetilde{u}_t^*)^k u_t^{p-k}.$$

Thus proceeding as with the proof of Corollary 1, by Proposition 2 and then Hölder's inequality, we conclude that (29) holds true if

$$E^* \widetilde{u}_t^{*p} - E\widetilde{u}_t^p = o_p(1). \tag{30}$$

Now, because both $\{\varepsilon_t^*\}_{t\in\mathbb{Z}}$ and $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ are *iid* sequences of random variables, we have that the left side of (30) is

$$E^*\left(\sum_{\ell=0}^{n+M} \vartheta_\ell \varepsilon_{t-\ell}^*\right)^p - E\left(\sum_{\ell=0}^{n+M} \vartheta_\ell \varepsilon_{t-\ell}\right)^p$$
$$= \sum_{\ell_1,\ldots,\ell_r=0}^{n+M} \left(\prod_{j=1}^{r} \vartheta_{\ell_j}^{q_j}\right) \left\{\prod_{j=1}^{r} E^*\left(\varepsilon_{t-\ell_j}^{*q_j}\right) - \prod_{j=1}^{r} E\left(\varepsilon_{t-\ell_j}^{q}\right)\right\},$$

where $\sum_{j=1}^{r} q_j = p$ and $q_j \geq 2$ for all $j = 1, \ldots, r$. Now we conclude by Corollary 1 and that $\{|\vartheta_\ell|^q\}_{\ell\geq 1}$ is a summable sequence when $q \geq 2$. ∎

Our next result shows that the bootstrap scheme is able to estimate correctly the covariance structure of the sequence $\{u_t\}_{t\in\mathbb{Z}}$.

PROPOSITION 3. *Under Conditions* 1 *and* 3*, we have that for* $0 \leq \ell \leq n$*,*

$$E^*\left(u_t^* u_{t+\ell}^*\right) - \gamma_u(\ell) = o_p(1)\,\ell^{2(d_1\wedge d_2)-1} + O\left(n^{2(d_1\wedge d_2)-1}\right).$$

PROOF. It is immediate by standard algebra using Lemmas 4, 9 and 10. ∎

We now verify in Propositions 4 and 5 that Bühlmann's (1997) Lemmas 5.4 and 5.5 hold true in our scenario.

PROPOSITION 4. *Assume Conditions 1 and 3 hold. Then, as $n \to \infty$, in probability,*

$$\varepsilon_t^* \overset{d}{\to}{}^* \varepsilon_t.$$

PROOF. Denote by $d_2(\cdot,\cdot)$ the Mallows metric as defined for example by Bickel and Freedman (1981). Let $\widehat{F}_n(x) = \frac{1}{n}\sum_{t=1}^n \mathcal{I}(\widehat{\varepsilon}_t \le x)$, $F_n(x) = \frac{1}{n}\sum_{t=1}^n \mathcal{I}(\varepsilon_t \le x)$ and $F(x) = P(\varepsilon_t \le x)$. Then

$$d_2\left(\widehat{F}_n, F\right) \le d_2\left(\widehat{F}_n, F_n\right) + d_2(F_n, F). \tag{31}$$

Let W be a random variable distributed uniformly on $\{1, 2, \ldots, n\}$. Then,

$$d_2\left(\widehat{F}_n, F_n\right) \le E_W(\varepsilon_W - \widehat{\varepsilon}_W)^2 = \frac{1}{n}\sum_{t=1}^n (\widehat{\varepsilon}_t - \varepsilon_t)^2.$$

By Proposition 1, the last expression converges to zero in probability. The second term of (31) converges to zero almost surely by Lemma 8.4 of Bickel and Freedman (1981). Therefore $d_2\left(\widehat{F}_n, F\right) = o_p(1)$ and the proposition holds. ∎

In the next step we extend Proposition 4 for the innovations ε_t^* to the observations u_t^*.

PROPOSITION 5. *Assume Conditions 1 and 3 hold. Then, as $n \to \infty$, in probability,*

$$u_t^* \overset{d}{\to}{}^* u_t.$$

COROLLARY 3. *Assume Conditions 1 and 3 hold. Then, for any finite collection $(t_1, ..., t_q)$, in probability,*

$$\left(u_{t_1}^*, .., u_{t_q}^*\right) \overset{d^*}{\Rightarrow} \left(u_{t_1}, .., u_{t_q}\right). \tag{32}$$

PROOF. This follows by Cramér-Wold device. It suffices to show that for any $c = (c_1, ..., c_q)'$, in probability,

$$\sum_{p=1}^q c_j u_{t_p}^* \overset{d^*}{\Rightarrow} \sum_{p=1}^q c_j u_{t_p}.$$

But proceeding as in Proposition 5, we conclude that (32) holds true. ∎

We finish this section giving some guidelines on how to choose m with real data sets. A first approach may be via cross-validation methods as in Beltrão and Bloomfield (1987), see also Robinson (1991). That is,

$$m = \arg\min_m \sum_{j=1}^{\tilde{n}} \left\{ \log \widehat{f}_u^-(m, \lambda_j) + \frac{I_{uu}(\lambda_j)}{\widehat{f}_u^-(m, \lambda_j)} \right\},$$

where $\widehat{f}_u^-(m, \lambda) = \frac{1}{2m}\sum_{k=-m;\neq 0}^m I_{uu}(\lambda_k - \lambda)$ is the leave-one-out average periodogram estimator.

A second approach is that employed in Lobato and Robinson (1997), see also Hidalgo (2008), where m is chosen according to

$$m(0) = \frac{1}{2}\left(\frac{3n}{4\pi}\right)^{3/4}\left|\frac{\gamma''(0)}{2\gamma(0)}\right|^{-3/8}$$

and $\gamma(\lambda)$ is the spectral density function of an $AR(1)$ sequence with parameter ρ, that is $\gamma(\lambda) = (2\pi)^{-1}\left(1+\rho^2 - 2\rho\cos(\lambda)\right)^{-1}$, although more general $\gamma(\lambda)$ functions can be adopted, see Lobato and Robinson (1997) for a discussion. Alternatively, we might employ

$$m = \frac{1}{2}\left(\frac{3n}{4\pi}\right)^{3/4}\frac{1}{\tilde{n}}\sum_{j=0}^{\tilde{n}}\left|\frac{\gamma''(\lambda_j)}{2\gamma(\lambda_j)}\right|^{-3/8} \tag{33}$$

which is, in a sense, the average pointwise bandwidths

$$m(\lambda_j) = 2^{-1}(3n/4\pi)^{3/4}\left|\gamma''(\lambda_j)/(2\gamma(\lambda_j))\right|^{3/8}.$$

In practice as ρ is not known, we would replace, say, $m(0)$ by

$$\begin{aligned} m^{**}(0) &= \frac{1}{2}\left(\frac{3n}{4\pi}\right)^{3/4}\left|\frac{\widehat{\gamma}''(0)}{2\widehat{\gamma}(0)}\right|^{-3/8} \\ &= \frac{1}{2}\left(\frac{3n}{4\pi}\right)^{3/4}\left|\frac{-\widehat{\rho}}{(1-\widehat{\rho})^2}\right|^{-3/8}, \end{aligned}$$

where $\widehat{\rho}$ is the least squares estimator of ρ. Because $m^{**}(0)$ could be smaller than 1 or greater than $\tilde{n}$, we truncate $m^{**}(0)$ as

$$m^{*} = \begin{cases} [\underline{m}] & \text{if } m^{**}(0) < [\underline{m}] \\ \left(\frac{3n}{4\pi}\right)^{3/4}\left|\frac{-\widehat{\rho}}{(1-\widehat{\rho})^2}\right|^{-3/8} & [\underline{m}] < m^{**}(0) < [\overline{m}] \\ [\overline{m}] & \text{if } [\overline{m}] < m^{**}(0), \end{cases}$$

where $\underline{m} = 0.06n^{3/4}$ and $\overline{m} = 1.2n^{3/4}$. Of course, similar comments can be used if one opts for the option of choosing m as in (33).

3. EXAMPLES OF THE VALIDITY OF THE BOOTSTRAP .

In this section we illustrate the validity of the bootstrap scheme for some examples or situations of interest in statistics. More specifically, we shall look at the least squares estimator in a time series regression model and model specification in the context of regression models.

3.1. ***Validity of the bootstrap in time series regression models***.

Consider the regression model

$$y_t = \alpha + \beta x_t + u_t, \quad t = 1, ..., n \tag{34}$$

and introduce the following condition on x_t.

CONDITION 4. *$\{x_t\}_{t\in\mathbb{Z}}$ is a sequence of random variables, mutually independent of $\{u_t\}_{t\in\mathbb{Z}}$, such that*

$$x_t = \sum_{j=0}^{\infty}\varphi_j\varrho_{t-j}; \quad \sum_{j=0}^{\infty}\varphi_j^2 < \infty, \quad a_0 = 1,$$

where $\{\varrho_t\}_{t\in\mathbb{Z}}$ *is a zero mean* iid *sequence of random variables with finite variance* σ_x^2 *and*

$$\varphi_j = \sum_{k=0}^{j} \overline{\varphi}_k c_{k-j},$$

where $\overline{\varphi}_k = \ell(k) k^{d_x - 1}$; $|\ell(k) - \ell(k+1)| < \ell'(k) k^{-1}$ *with* $\ell'(k) > 0$, $d_x \in \left[0, \frac{1}{2}\right)$ *and* $\sum_{k=0}^{\infty} k^2 |c_k| < \infty$.

Condition 4 allows for the sequence $\{x_t\}_{t\in\mathbb{Z}}$ to exhibit LM dependence. Following the results in Robinson and Hidalgo (1997), we can weaken this condition as their results do not require x_t to be a linear sequence at all, i.e. (34) can be modified to $y_t = \alpha + \beta\zeta(x_t) + u_t$. However, since our purpose is to illustrate the validity of the bootstrap scheme in this scenario, we keep it for notational convenience. For the same reason, we shall modify our Condition 1 in that we shall assume that $d_2 = 0$. It is obvious that the results would not be affected if d_2 were greater than zero, but it would not add anything substantial or relevant. That is,

CONDITION 5. *Condition* 1 *holds except that* $d_2 = 0$.

Denote the least squares estimator, LSE, of β as

$$\widehat{\beta} = \left(\sum_{t=1}^{n} (x_t - \overline{x})^2\right)^{-1} \sum_{t=1}^{n} (x_t - \overline{x}) y_t,$$

where $\overline{z} = n^{-1}\sum_{t=1}^{n} z_t$ for a generic sequence $\{z_t\}_{t=1}^{n}$. We first examine our estimator of d_1.

Compute the residuals as

$$\widehat{u}_t = y_t - \overline{y} - \widehat{\beta}(x_t - \overline{x}); \quad t = 1, ..., n. \tag{35}$$

We then estimate d_1 as in (14) but with u_t replaced there by $\widehat{u}_t$, that is

$$\widehat{d}_1 = \arg\min_{d\in[0,\Delta]} \left\{ \log\left(\frac{1}{m}\sum_{j=1}^{m} \lambda_j^{2d} I_{\widehat{u}\widehat{u}}(\lambda_j)\right) - \frac{2d}{m}\sum_{j=1}^{m} \log\lambda_j \right\}, \tag{36}$$

where $I_{\widehat{u}\widehat{u}}(\lambda)$ is the periodogram of $\{\widehat{u}_t\}_{t=1}^{n}$. Similarly the estimator of $h(\lambda)$ becomes

$$\widehat{h}(\lambda) = \frac{1}{2m+1}\sum_{k=-m}^{m} \left|1 - e^{-i(\lambda+\lambda_k)}\right|^{2\widehat{d}_1} \frac{I_{\widehat{u}\widehat{u}}(\lambda+\lambda_k)}{2\pi}. \tag{37}$$

We have now the following proposition.

PROPOSITION 6. *Under Conditions* 3,4 *and* 5, *if* $d_1 + d_x < 1/2$, *as* $n \to \infty$,

$$\widehat{d}_1 - d_1 = O_p\left(m^{-1/2}\right).$$

PROOF. Similar to Robinson (1997), the properties of the estimator are not affected by using the residuals $\widehat{u}_t$ instead of the true errors u_t. Indeed, by definition,

$$I_{\widehat{u}\widehat{u}}(\lambda_j) - I_{uu}(\lambda_j) = \left(\widehat{\beta} - \beta\right) I_{ux}(\lambda_j) + \left(\widehat{\beta} - \beta\right)^2 I_{xx}(\lambda_j),$$

$\widehat{\beta}-\beta=O_p\left(n^{-1/2}\right)$ and proceeding as in Robinson $(1995a)$, we have that

$$E\left|I_{ux}\left(\lambda_j\right)I_{ux}\left(\lambda_k\right)\right|=E\left(w_u\left(\lambda_j\right)\overline{w}_u\left(\lambda_k\right)\right)E\left(w_x\left(\lambda_j\right)\overline{w}_x\left(\lambda_k\right)\right)$$
$$\simeq\lambda_j^{-d_1-d_x}\lambda_k^{-d_1-d_x}\max\left(k^{-1},j^{-1}\right).$$

So, we easily conclude that

$$\frac{1}{m}\sum_{j=1}^{m}\lambda_j^{2d_1}\left(I_{\widehat{u}\widehat{u}}\left(\lambda_j\right)-I_{uu}\left(\lambda_j\right)\right)=o_p\left(m^{-1/2}\right).$$

From here the proof proceeds as in Robinson (1997). ∎

We now describe the bootstrap in the following *4 STEPS*, where the first *3 STEPS* are an obvious reformulation of *STEPS 1* to *3* given in Section 2.

STEP 1 We compute the innovations as

$$\widehat{\varepsilon}_t=\frac{1}{n^{1/2}}\sum_{j=1}^{n-1}e^{it\lambda_j}\widehat{\Phi}\left(\lambda_j\right)w_{\widehat{u}}\left(\lambda_j\right),\quad t=1,...,n,$$

where $\widehat{\Phi}\left(\lambda\right)=g_1\left(\lambda,\widehat{d}_1;n-M\right)\widehat{A}\left(\lambda\right)$, $g_1\left(\lambda,d_1;L\right)$ given in (13) and $\widehat{A}\left(\lambda\right)$ is computed similarly as in *STEP 1* but with $\widehat{h}\left(\lambda_j\right)$ given in (37) instead of (16) in the definition of $\widehat{C}\left(\lambda\right)$ given in (17).

STEPS 2 and 3 As those in Section 2.

STEP 4 Construct the bootstrap sample y_t^* as

$$y_t^*=\overline{y}+\widehat{\beta}\left(x_t-\overline{x}\right)+u_t^*,\quad t=1,...,n. \tag{38}$$

Compute the bootstrap LSE $\widehat{\beta}^*$ as

$$\widehat{\beta}^*=\left(\sum_{t=1}^{n}\left(x_t-\overline{x}\right)^2\right)^{-1}\sum_{t=1}^{n}\left(x_t-\overline{x}\right)y_t^*, \tag{39}$$

being $\widehat{u}_t^*=y_t^*-\overline{y}^*+\widehat{\beta}^*\left(x_t-\overline{x}\right)$ the least squares residuals.

REMARK 5. *One of the motivations to keep x_t fixed in the bootstrap algorithm comes from results/observation in Horowitz* (1997)*, who shows that there is no advantage by "bootstrapping" also the regressor x_t.*

Denote the spectral density function of x_t by $f_x\left(\lambda\right)$.

PROPOSITION 7. *Assuming Conditions* 3,4 *and* 5*, if $d_1+d_x<1/2$, we have that,* in probability,

$$n^{1/2}\left(\widehat{\beta}^*-\widehat{\beta}\right)\overset{\mathcal{L}}{\to}\mathcal{N}\left(0,\mathcal{V}\right),$$

where $\mathcal{V}=\left(\sigma_x^2\right)^{-2}\int_{-\pi}^{\pi}f_u\left(\lambda\right)f_x\left(\lambda\right)d\lambda$ is the asymptotic variance of the LSE.

Looking at the proofs of Proposition 7 and Robinson and Hidalgo's (1997) Theorems 1 and 2, we envisage that Proposition 7 holds true for the bootstrap analogue of the estimator of β proposed in the latter manuscript. That is, let the estimator of β be

$$\widehat{\beta}_\psi=\frac{\sum_{t,s=1}^{n}\psi_{t-s}\left(x_t-\overline{x}\right)\left(y_s-\overline{y}\right)}{\sum_{t,s=1}^{n}\psi_{t-1}\left(x_t-\overline{x}\right)\left(x_s-\overline{x}\right)}, \tag{40}$$

where

$$\psi_\ell = \frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\Gamma(\lambda)\cos\ell\lambda d\lambda$$

and the function $\Gamma(\lambda)$ satisfies $\Delta = \int_{-\pi}^{\pi}\Gamma^2(\lambda) f_u(\lambda) f_x(\lambda)\, d\lambda < \infty$. Then, the bootstrap analogue of $\widehat{\beta}_\psi$, defined as

$$\widehat{\beta}^*_\psi = \frac{\sum_{t,s=1}^{n}\psi_{t-s}(x_t-\overline{x})(y^*_s-\overline{y}^*)}{\sum_{t,s=1}^{n}\psi_{t-s}(x_t-\overline{x})(x_s-\overline{x})},$$

will be valid for $\widehat{\beta}_\psi$. Notice that when $\Gamma(\lambda)=1$, $\widehat{\beta}_\psi$ becomes the LSE in (34), whereas the generalized least squares, GLS, estimator is obtained when $\Gamma(\lambda) = f_u^{-1}(\lambda)$. In addition, we would not require the assumption that $d_1 + d_x < 1/2$, neither for the results of Proposition 6 nor for the validity of $\widehat{\beta}^*_\psi$ as

$$n^{1/2}\left(\widehat{\beta}_\psi - \beta\right) \xrightarrow{d} \mathcal{N}\left(0, \left(\int_{-\pi}^{\pi}\Gamma(\lambda) f_x(\lambda)\, d\lambda\right)^{-2}\Delta\right)$$

for any $0 \le d_x, d_1 < 1/2$.

3.2. ***Validity of the bootstrap for model specification*.**

Because the purpose of this section is to illustrate the validity of our bootstrap scheme, we shall only consider the case where the sequence x_t is Gaussian and consider the question of the correct specification of (34). For a more comprehensive set of results which include when x_t is nonGaussian and/or nonlinear regression models and linear regression models with no intercept α, we refer to Hidalgo (2019). The main reason is because the asymptotic distribution of $\mathcal{T}_n(x)$ in (42) below depends, among other issues, on whether $\{x_t\}_{t\in\mathbb{Z}}$ is or is not a Gaussian sequence as Koul, Baillie and Surgailis (2004) and Hidalgo (2019) have showed. Thus, we consider the hypothesis testing

$$H_0: \ E[y_t \mid x_t] = \alpha + \beta x_t \tag{41}$$

being our alternative hypothesis H_1 the negation of the null. Following Stute (1997), the testing procedure will be based on the partial sums empirical process

$$\mathcal{T}_n(x) = \frac{1}{n}\sum_{t=1}^{n}\mathcal{I}(x_t < x)\,\widehat{u}_t, \tag{42}$$

where $\widehat{u}_t$ was given in (35) and $\mathcal{I}(\circ)$ denotes the indicator function. The bootstrap scheme is given in the next *5 STEPS*.

STEP 1-4 As given in the previous section.
STEP 5 Compute the bootstrap analogue of $\mathcal{T}_n(x)$ as

$$\mathcal{T}^*_n(x) = \frac{1}{n}\sum_{t=1}^{n}\mathcal{I}(x_t < x)\,\widehat{u}^*_t.$$

We need to introduce an extra condition for the validity of our results.

CONDITION 6. *$\{\varrho_t\}_{t\in\mathbb{Z}}$ and $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ are* iid *sequences of random variables with finite 8th moments. In addition, denoting $\phi(x)$ as the probability density function of x_t, we have that*

$$\int_{\mathbb{R}}\sum_{p=1}^{4}\left|\frac{\partial^p\phi}{\partial x^p}(x)\right| dx < \infty.$$

Condition 6 is very mild and similar to that in Wu (2003). Although when x_t is Gaussian, the condition is redundant, we keep it because the results of Theorem 1 holds true regardless whether x_t is or it is not Gaussian provided that Condition 6 holds. We now introduce some notation.

We shall write

$$\mathring{\mathbf{i}}_t(x) = \mathcal{I}(x_t < x) - F(x) - G(x)\, x_t, \tag{43}$$

where $F(x) = \int_{-\infty}^{x} \phi(z)\, dz$ and $G(x) = E(\mathcal{I}(x_t < x)\, x_t)$.

Some comments regarding $\mathring{\mathbf{i}}_t(x)$ are relevant and helpful to understand why the results that follow are different when $\{x_t\}_{t\in\mathbb{Z}}$ is nonGaussian. When $\{x_t\}_{t\in\mathbb{Z}}$ is Gaussian, $G(x) = -\phi(x)$, so that $F(x) - \phi(x)\, x_t$ becomes the first two terms in the Hermite expansion of $\mathcal{I}(x_t < x)$, see Dehling and Taqqu (1989), whereas for non-Gaussian linear sequences, it becomes the first two terms of the expansion of $\mathcal{I}(x_t < x)$ in terms of its Appell expansion, see Giraitis and Surgailis (1994). It is well known that under Gaussianity $E\left(\mathring{\mathbf{i}}_t(x)\, x_t\right) = 0$, whereas the latter is not guaranteed if x_t is nonGaussian, due to the lack of orthogonality of the Appell polynomials. This rather innocuous result plays a key and pivotal role when examining the statistical properties of $\mathcal{T}_n(x)$, see Koul et al. (2004) or Hidalgo (2019) for some details. For the sake of convenience we shall state a few results regarding the statistical behaviour of $\mathcal{T}_n(x)$. For a proof we refer to the aforementioned manuscripts of Koul et al. (2004) or Hidalgo (2019).

We denote $\mathcal{G}(x)$ a Gaussian process in $x \in \mathbb{R}$ with *covariance* structure given by

$$Cov(\mathcal{G}(x), \mathcal{G}(y)) = \gamma_u(0)\, E\left(\mathring{\mathbf{i}}_0(x)\, \mathring{\mathbf{i}}_0(y)\right) + \sum_{\ell=1}^{\infty} \gamma_u(\ell)\, E\left(\mathring{\mathbf{i}}_0(x)\, \mathring{\mathbf{i}}_\ell(y)\right)$$
$$+ \sum_{\ell=1}^{\infty} \gamma_u(\ell)\, E\left(\mathring{\mathbf{i}}_0(y)\, \mathring{\mathbf{i}}_\ell(x)\right).$$

Observe that because $E\left(\mathring{\mathbf{i}}_\ell(x)\, \mathring{\mathbf{i}}_0(x)\right) = O\left(\ell^{4d_x - 2}\right)$, see Wu (2003) or Dehling and Taqqu (1989) under Gaussianity, and Condition 5 implies that $\gamma_u(\ell) = O\left(\ell^{2d_1 - 1}\right)$, we conclude that $|Cov(\mathcal{G}(x), \mathcal{G}(y))| < C$ for all $x, y \in \mathbb{R}$ if $d_1 + 2d_x < 1$. It is worth mentioning that, in view of our comments at the end of last section, under the latter condition, the LSE might not be asymptotically Gaussian neither $n^{1/2}-$consistent.

THEOREM 1. *Assume Conditions* 4,5 *and* 6. *Then, if* $d_1 + 2d_x < 1$, *we have that*

$$\mathcal{G}_n(x) := \frac{1}{n^{1/2}} \sum_{t=1}^{n} \mathring{\mathbf{i}}_t(x)\, u_t \overset{weakly}{\Rightarrow} \mathcal{G}(x) \quad x \in \mathbb{R}.$$

The results of Theorem 1 are valid under either LM or SM dependence. However it is more relevant to notice that the results hold true either $\{x_t\}_{t\in\mathbb{Z}}$ is a sequence of Gaussian random variables or not. We now look at the behaviour of $\mathcal{T}_n(x)$, which is a consequence of Theorem 1.

PROPOSITION 8. *Assume Conditions* $3, 4$,5 *and* 6 *with* $\{x_t\}_{t\in\mathbb{Z}}$ *being a sequence of Gaussian random variables. Then under* H_0, *we have that, if* $d_1 + 2d_x < 1$, *we have that uniformly in* $x \in \mathbb{R}$,

$$n^{1/2}\mathcal{T}_n(x) = \mathcal{G}_n(x) + o_p(1) \overset{weakly}{\Longrightarrow} \mathcal{G}(x).$$

COROLLARY 4. *Under the conditions of Proposition 8, we have that for any continuous functional* $\varphi(\cdot):\mathbb{R}\to\mathbb{R}^{+}$*, if* $d_1+2d_x<1$*,*

$$\varphi\left(n^{1/2}\mathcal{T}_n(x)\right)\overset{d}{\to}\varphi(\mathcal{G}(x)).$$

PROOF. The proof is standard by the continuous mapping theorem and Proposition 8, so it is omitted. ∎

Standard functionals $\varphi(\cdot)$ are the Kolmogorov-Smirnov and the Cramér von Mises. The former is the $\mathcal{L}_\infty$-norm whereas the latter is the $\mathcal{L}_2$-norm, and they are given respectively by

$$\mathcal{KS}_n=\max_{\ell=1,\dots,N}\left|n^{1/2}\mathcal{T}_n(x_\ell)\right|;\quad \mathcal{C}v\mathcal{M}_n=\frac{1}{N}\sum_{\ell=1}^{N}\left|n^{1/2}\mathcal{T}_n(x_\ell)\right|^2,$$

where $\{x_\ell\}_{\ell=1}^{N}$ forms a set dense in any compact set $\mathcal{X}\subset\mathbb{R}$.

THEOREM 2. *Assuming Conditions* $3,4,5$ *and* 6*, if* $d_1+2d_x<1$*, we have that,* in probability,

$$\mathcal{G}_n^{*}(x):=\frac{1}{n^{1/2}}\sum_{t=1}^{n}\mathring{\mathbf{1}}_t(x)\,u_t^{*}\overset{d^*}{\Rightarrow}\mathcal{G}(x)\qquad x\in\mathbb{R}.$$

It is important to notice that the results of Theorem 2 hold true either for $\{x_t\}_{t\in\mathbb{Z}}$ being a Gaussian sequence or not. The only condition that we used there was that the x_t is a linear sequence and the order of magnitude of the covariance of the sequence $\left\{\mathring{\mathbf{1}}_t(x)\right\}_{t\in\mathbb{Z}}$. We now examine the behaviour of $\mathcal{T}_n^{*}(x)$.

PROPOSITION 9. *Assuming Conditions* $3,4,5$ *and* 6 *with* $\{x_t\}_{t\in\mathbb{Z}}$ *being a sequence of Gaussian random variables, we have that,* in probability,

$$\text{(a)}\quad n^{1/2}\mathcal{T}_n^{*}(x)\overset{d^*}{\Rightarrow}\mathcal{G}(x)\quad x\in\mathbb{R}$$

$$\text{(b)}\quad \varphi\left(n^{1/2}\mathcal{T}_n^{*}(x)\right)\overset{\mathcal{L}}{\to}\varphi(\mathcal{G}(x))$$

for any continuous functional $\varphi(\cdot):\mathbb{R}\to\mathbb{R}^{+}$.

4. MONTE CARLO EXPERIMENT. We present a Monte Carlo experiment to shed some light on the behaviour of the bootstrap scheme for the least squares and also our test. A more complete set of scenarios can be obtained in the Appendix of the supplementary material, where we consider even when $2d_x+d_1>1$ or $d_x+d_1>1/2$ and x_t is a non-Gaussian sequence.

To address the performance of the bootstrap least squares and the test under the null hypothesis, we have generated the linear regression model

$$y_t=\alpha+\beta x_t+u_t,\ t=1,\dots,n, \tag{44}$$

where $\alpha=\beta=1$ for two different sample sizes $n=128$ and 512. When we examine the performance of the bootstrap for the least squares estimator of β, we have considered $d_x=0.1,0.2,0.3$ and $d_1=0.2,0.3$. Also to shed some light as how the proposed bootstrap compares to the sieve bootstrap, we have considered the case when the errors follow an $AR(1)$ or $MA(1)$ sequence with parameters 0.7 and 0.9. On the other hand, when addressing the performance of the test, we considered the scenarios $d_x=0.1,..,0.4$ and $d_1=0.2$.

In the latter scenario we also present the results when the innovations of the regressor were a χ_1^2 centered around its mean. All throughout the errors $\{u_t\}_{t=1}^n$ were generated as a sequence of Gaussian random variables with mean 0. The statistic $\mathcal{T}_n(x)$ were computed in the range $x \in [-1.0, 1.0]$ with a mesh width of 0.1 and we have chosen the *Kolmogorov's* type of functional for $\varphi(\cdot)$. That is,

$$\mathcal{KS}_n = \max_{\ell=1,\ldots,21} \left| n^{1/2} \mathcal{T}_n(x_\ell) \right|,$$

where $\{x_\ell\}_{\ell=1}^{21}$, $x_\ell = -1.0 + (\ell - 1)\, 0.1$.

In order to save computational time, for each sample we compute only one bootstrap counterpart according to Section 3 and equations (3.1) and (3.2). The stacked bootstrapped statistics are then used to construct critical values and confidence regions at appropriate levels. For each combination of models and/or samples sizes n, 1000 iterations were performed. This is the idea behind the WARP algorithm of Giacomini et al. (2013). Finally, to implement the bootstrap algorithm we need to choose the smoothing parameter m. Although an algorithm as that described at then end of the previous section can be implemented, in this Monte-Carlo experiment we have considered two different choices of m, namely $m = n/4$ and $m = n/8$. Likewise in the expression $\widehat{C}(\lambda) = \exp\left\{\sum_{r=1}^{[n/4m]} \widehat{c}_r e^{-ir\lambda}\right\}$, we have chosen $\widehat{c}_r = 0$ for $r \geq 1$ and the case $\widehat{c}_r = 0$ for $r > 1$ with $\widehat{c}_1 \neq 0$. The first scenario uses the fact that we know that there is no SM component whereas in the second we have taken $[n/4m] = 1$, after we notice that in almost all cases $[n/4m] \leq 1$. Finally, in all the tables, the first row in each cell presents the results of the test for the 10% size whereas the second row are those for the 5% size.

Table 1 presents the results of the size when testing $\beta = 0$ against the alternative that it is different than zero using the critical values from the asymptotic Gaussian random variable or those obtained under the bootstrap scheme, whereas Table 2 presents the results when u_t is weak dependent.

TABLE 1

Size using the asymptotic critical values

d_1	.2		.2		.3		.3	
	n=128		n=512		n=128		n=512	
d_x	m= $\frac{n}{4}$	$\frac{n}{8}$	$\frac{n}{4}$	$\frac{n}{8}$	$\frac{n}{4}$	$\frac{n}{8}$	$\frac{n}{4}$	$\frac{n}{8}$
.1	20.6	19.9	21.4	21.4	20.4	20.4	21.1	21.1
	10.5	10.4	9.6	9.6	10.8	9.7	10.5	10.5
.2	23.1	21.9	22.1	22.1	21.9	22.9	21.6	21.6
	11.7	12.3	10.7	10.7	11.8	11.8	10.0	10.0
.3	22.7	21.3	21.8	21.8	26.8	25.0	23.1	23.1
	12.4	12.0	12.0	12.0	14.1	13.5	12.4	12.4

Size using the bootstrap critical values

d_1	.2		.2		.3		.3	
	n=128		n=512		n=128		n=512	
d_x	m= $\frac{n}{4}$	$\frac{n}{8}$	$\frac{n}{4}$	$\frac{n}{8}$	$\frac{n}{4}$	$\frac{n}{8}$	$\frac{n}{4}$	$\frac{n}{8}$
.1	10.1	9.5	9.1	9.1	8.6	9.7	10.2	10.0
	4.3	5.6	5.1	5.2	4.7	4.8	4.7	4.5
.2	12.3	10.8	9.9	10.1	9.6	10.0	11.7	11.6
	7.1	4.8	5.2	5.2	4.8	5.3	6.6	6.7
.3	9.4	12.0	11.4	11.3	12.4	12.2	10.9	10.8
	4.9	5.8	6.0	6.6	6.9	5.8	4.8	5.9

TABLE 2

Size using the bootstrap critical values

$AR(1)$	n=128				n=512			
$\rho = .7$	$m = \frac{n}{4}$	$m = \frac{n}{8}$	AR(1)	AR(3)	$m = \frac{n}{4}$	$m = \frac{n}{8}$	AR(1)	AR(3)
$d_x = .1$	5.0	4.9	4.2	4.7	4.3	3.7	4.6	4.6
	9.5	9.5	7.6	8.8	8.5	8.7	8.3	9.5
$d_x = .2$	4.8	4.8	5.0	6.5	7.5	7.6	3.8	4.2
	11.1	11.3	12.6	13.7	12.0	12.0	11.1	9.1
$d_x = .3$	5.0	4.7	6.8	5.8	4.7	4.6	6.1	6.2
	11.7	11.2	13.1	10.0	9.3	9.0	11.3	10.8
$AR(1)$	n=128				n=512			
$\rho = .9$	$m = \frac{n}{4}$	$m = \frac{n}{8}$	AR(1)	AR(3)	$m = \frac{n}{4}$	$m = \frac{n}{8}$	AR(1)	AR(3)
$d_x = .1$	3.7	4.0	6.2	6.0	6.6	6.6	4.4	3.0
	8.5	8.3	10.9	11.6	13.1	12.5	9.9	9.3
$d_x = .2$	4.1	4.6	6.1	5.4	5.7	5.4	5.4	4.6
	9.5	9.4	11.3	10.2	9.6	9.7	11.0	10.4
$d_x = .3$	4.0	4.2	3.9	4.8	5.1	5.0	5.2	5.0
	9.4	9.6	9.1	12.2	9.6	9.6	9.2	10.8
$MA(1)$	n=128				n=512			
$\theta = .7$	$m = \frac{n}{4}$	$m = \frac{n}{8}$	AR(1)	AR(3)	$m = \frac{n}{4}$	$m = \frac{n}{8}$	AR(1)	AR(3)
$d_x = .1$	5.2	5.2	6.5	4.0	4.4	4.9	4.8	5.8
	10.6	10.1	10.6	9.8	8.5	9.0	9.8	11.8
$d_x = .2$	5.7	5.8	5.2	4.8	4.7	4.7	4.9	4.7
	8.7	9.2	9.9	10.5	10.1	10.2	9.0	9.1
$d_x = .3$	4.6	4.7	4.6	5.5	4.0	3.9	3.5	4.8
	9.7	9.2	10.1	10.2	9.7	10.1	8.5	9.5
$MA(1)$	n=128				n=512			
$\theta = .9$	$m = \frac{n}{4}$	$m = \frac{n}{8}$	AR(1)	AR(3)	$m = \frac{n}{4}$	$m = \frac{n}{8}$	AR(1)	AR(3)
$d_x = .1$	7.6	8.4	5.1	5.2	3.9	4.0	3.9	4.5
	13.5	14.0	11.3	12.0	8.4	9.5	9.1	10.0
$d_x = .2$	5.0	5.2	3.7	4.4	4.6	4.4	6.3	6.2
	10.3	10.1	10.2	9.6	10.0	10.4	10.1	7.8
$d_x = .3$	4.0	4.3	4.5	5.4	4.5	4.5	6.5	6 .2
	8.9	9.5	10.6	11.4	9.7	9.5	11.1	10.4

A general conclusion from *Table 1* is the good performance of the bootstrap scheme even for samples sizes as small as $n = 128$, and that it gives a big improvement when compared the size obtained using the asymptotic critical values from the standard Gaussian random variable. Also, even when the LSE is known not to be Gaussian, i.e. when $d_u = d_x = 0.3$, the bootstrap scheme appears to approximate the finite sample distribution quite well. On the other hand, *Table 2* suggests that our proposed bootstrap compares favourably to the sieve bootstrap.

TABLE 3

Size with $\widehat{C}(\lambda) = 1$

x_t	Normal		Normal		χ_2^2		χ_2^2	
	n=128		n=512		n=128		n=512	
d_x	m= $\frac{n}{4}$	$\frac{n}{8}$	$\frac{n}{4}$	$\frac{n}{8}$	$\frac{n}{4}$	$\frac{n}{8}$	$\frac{n}{4}$	$\frac{n}{8}$
.1	7.8	6.2	10.5	7.8	9.3	7.4	11.4	8.0
	4.0	3.5	5.2	4.0	5.0	3.1	3.7	2.9
.2	10.1	9.8	11.9	10.6	9.0	9.9	9.9	8.9
	4.8	4.9	5.8	5.9	3.7	4.7	6.2	3.7
.4	9.3	7.0	9.7	8.0	11.2	9.1	10.0	10.5
	4.5	3.1	4.7	5.1	4.7	4.2	5.8	5.0

Size with $\widehat{C}(\lambda) = \exp\left\{\widehat{c}_1 e^{-ir\lambda}\right\}$

x_t	Normal		Normal		χ_2^2		χ_2^2	
	n=128		n=512		n=128		n=512	
d_x	m= $\frac{n}{4}$	$\frac{n}{8}$	$\frac{n}{4}$	$\frac{n}{8}$	$\frac{n}{4}$	$\frac{n}{8}$	$\frac{n}{4}$	$\frac{n}{8}$
.1	12.5	9.8	11.1	9.1	10.6	11.4	9.7	8.8
	6.1	6.1	6.9	3.5	6.2	5.7	4.2	6.5
.2	11.1	8.0	10.2	8.5	9.9	8.6	9.3	11.7
	4.7	4.2	4.9	4.5	5.1	4.3	5.2	5.2
.4	12.4	11.7	10.7	12.8	10.9	8.5	12.7	11.7
	6.8	5.5	4.5	7.3	5.1	3.4	7.5	5.8

A general conclusion from *Table 3* is the good performance of the test even for samples sizes as small as $n = 128$. This performance is regardless the distribution of the regressor x_t and the choice of $m = n/4$ appears to perform slightly better for moderate sample sizes, i.e. when $n = 128$. Also the tables suggests that even when we choose $\widehat{c}_1 \neq 0$, there is no visible deterioration of the finite sample performance when compared to the case of $\widehat{c}_1 = 0$.

To address the power of the test we simulated the regression model

$$y_t = \alpha + \beta x_t + \gamma \sin(x_t) + u_t, \; t = 1, ..., n \tag{45}$$

with $\gamma = 0.5$ and 1.5. We present the results of the Monte Carlo experiment in *Table 4* below with $\widehat{C}(\lambda) = \exp\left\{\widehat{c}_1 e^{-ir\lambda}\right\}$.

TABLE 4

Power for model with $\gamma = 0.5$

x_t	Normal		Normal		χ^2_2		χ^2_2	
	n=128		n=512		n=128		n=512	
d_x	m= $\frac{n}{4}$	$\frac{n}{8}$	$\frac{n}{4}$	$\frac{n}{8}$	$\frac{n}{4}$	$\frac{n}{8}$	$\frac{n}{4}$	$\frac{n}{8}$
.1	19.9	17.4	53.8	52.9	77.6	67.2	99.9	100
	8.5	8.7	43.4	40.0	64.4	52.4	99.8	99.8
.2	23.5	17.3	59.3	61.2	73.1	65.6	100	99.9
	13.5	8.2	42.1	45.3	58.7	53.4	99.4	99.6
.4	37.3	34.0	85.9	85.5	72.0	71.6	100	100
	26.9	21.6	73.7	77.7	58.8	56.0	100	100

Power for model with $\gamma = 1.5$

x_t	Normal		Normal		χ^2_2		χ^2_2	
	n=128		n=512		n=128		n=512	
d_x	m= $\frac{n}{4}$	$\frac{n}{8}$	$\frac{n}{4}$	$\frac{n}{8}$	$\frac{n}{4}$	$\frac{n}{8}$	$\frac{n}{4}$	$\frac{n}{8}$
.1	74.8	69.9	99.7	100	100	100	100	100
	61.8	57.2	99.6	99.8	100	100	100	100
.2	79.0	75.9	100	100	100	100	100	100
	69.4	60.1	100	99.9	100	100	100	100
.4	89.9	88.1	100	100	100	100	100	100
	82.6	80.7	100	100	100	100	100	100

The results for (45) illustrate a very good power performance of the test, and as expected it improves as the value of γ or the sample size increases.

5. CONCLUSIONS. This paper has introduced a bootstrap scheme in the frequency domain which is valid when the data exhibits either SM or LM dependence. The bootstrap has some similarities with the TFT given in Kirsch and Politis (2011) in that we obtain time series resamples. On the other hand, the scheme is similar to the $AR-$sieve bootstrap in that it is able to match the moments of the data correctly. We have also illustrated the validity of the scheme in some situations/statistics of interest. Namely for the LSE and model specification in a time series regression model context.

Acknowledgments. I thank the Associate Editor and three referees for helpful comments which led to a much improved and clearer version of the article. Also I thank Hao Dong and Chen Qiu for their excellent Monte Carlo computations. Of course, all remaining errors are my sole responsibility.

SUPPLEMENTARY MATERIAL

Supplement: Bootstrap Long Memory Processes in the Frequency Domain (http://www.e-publications.org/ims/support/dowload/imsart-ims.zip). All the technical details and Tables are provided in the Supplementary Section.

6. PROPOSITIONS. In what follows, K denotes a generic finite and positive constant. We shall give the proofs of our main results

6.0.1. ***Proof of Proposition 1***.
First observe that because

$$\frac{1}{n}\sum_{t=1}^{n}\left|\check{\varepsilon}_t-\varepsilon_t\right|^p \leq 2^{p-1}\left(\frac{1}{n}\sum_{t=1}^{n}\left|\widehat{\varepsilon}_t-\varepsilon_t\right|^p+\left|\frac{1}{n}\sum_{t=1}^{n}\widehat{\varepsilon}_t\right|^p\right),$$

it suffices to show that the first term on the right is $o_p(1)$.

We shall first examine the case when $B(L)=1$. Denoting

$$\widetilde{\varepsilon}_t=\frac{1}{n^{1/2}}\sum_{j=1}^{n}e^{it\lambda_j}\left(\sum_{\ell=0}^{n}\xi_\ell e^{-i\ell\lambda_j}\right)w_u(\lambda_j),$$

and observing that when $B(L)=A(L)=1$, $\widehat{\varepsilon}_t$ in *STEP 1* becomes

$$\widehat{\varepsilon}_t=\frac{1}{n^{1/2}}\sum_{j=1}^{n-1}e^{it\lambda_j}\left(\sum_{\ell=0}^{n}\widehat{\xi}_\ell e^{-i\ell\lambda_j}\right)w_u(\lambda_j),\quad t=1,...,n,$$

we then have that (21) and Lemma 3 imply that

$$\begin{aligned}\widehat{\varepsilon}_t-\widetilde{\varepsilon}_t&=\sum_{\ell=0}^{t-1}\left(\widehat{\xi}_\ell-\xi_\ell\right)u_{t-\ell}+\sum_{\ell=t}^{n}\left(\widehat{\xi}_\ell-\xi_\ell\right)u_{n-(\ell-t)}\\&=\sum_{h=1}^{H-1}\left\|\widehat{d}-d\right\|^h\sum_{\ell=0}^{t-1}g_\ell^h(d)\,\xi_\ell u_{t-\ell}+\log^H n\left\|\widehat{d}-d\right\|^H\sum_{\ell=0}^{t-1}\left|\frac{u_{t-\ell}}{\ell^{1+(d_1\wedge d_2)}}\right|\\&\quad+\sum_{h=1}^{H-1}\left\|\widehat{d}-d\right\|^h\sum_{\ell=t}^{n}g_\ell^h(d)\,\xi_\ell u_{n-(\ell-t)}+\log^H n\left\|\widehat{d}-d\right\|^H\sum_{\ell=t}^{n}\left|\frac{u_{n-(\ell-t)}}{\ell^{1+(d_1\wedge d_2)}}\right|.\end{aligned}$$

By standard inequalities, it suffices to show that

$$\frac{1}{n}\sum_{t=1}^{n}\left|\widehat{\varepsilon}_t-\widetilde{\varepsilon}_t\right|^p=o_p(1) \tag{46}$$

$$\frac{1}{n}\sum_{t=1}^{n}\left|\varepsilon_t-\widetilde{\varepsilon}_t\right|^p=o_p(1). \tag{47}$$

That (47) holds true follows because $\varepsilon_t-\widetilde{\varepsilon}_t=:\sum_{\ell=t}^{\infty}\xi_\ell u_{t-\ell}-\sum_{\ell=t}^{n}\xi_\ell u_{n-(\ell-t)}$ so that the expectation of the left side of (47) is bounded by

$$\frac{K}{n}\sum_{t=1}^{n}\left(\sum_{\ell=t}^{\infty}|\xi_k|\right)^p\leq\frac{K}{n}\sum_{t=1}^{n}\frac{\mathcal{I}(d_1\wedge d_2>0)}{t^{p(d_1\wedge d_2)}}=o(1),$$

because $|\xi_k|=O\left(k^{-1-(d_1\wedge d_2)}\mathcal{I}(d_1\wedge d_2>0)\right)$, that $E|u_t|^p<K$ and Hölder's inequality yields that

$$E\left|\sum_{k=a}^{b}\zeta_k\eta_k\right|^p\leq\left(\sum_{k=a}^{b}|\zeta_k|\right)^{p-1}\sum_{k=a}^{b}|\zeta_k|\,E|\eta_k|^p. \tag{48}$$

Next (46) also holds true after we observe that (48) implies that

$$E\left|\sum_{\ell=0}^{t-1}g_\ell^h(d)\,\xi_\ell u_{t-\ell}\right|^p+E\left|\sum_{\ell=t}^{n}g_\ell^h(d)\,\xi_\ell u_{n-(t-\ell)}\right|^p\leq K\sum_{\ell=0}^{n}\left|g_\ell^h(d)\,\xi_\ell\right|\leq K,$$

$\left\|\widehat{d}-d\right\| = O_p\left(m^{-1/2}\right)$ and choosing H large enough in Lemma 3.

Now, we examine the general case when $B(L) \neq 1$. To that end, denote

$$\mathring{\varepsilon}_t = \sum_{\ell=0}^{t-1} \mathring{\phi}_\ell u_{t-\ell} + \sum_{\ell=t}^{n} \mathring{\phi}_\ell u_{n-(\ell-t)},$$

where

$$\mathring{\phi}_\ell = \begin{cases} \sum_{k=0}^{\ell \wedge M} \xi_{\ell-k} a_k, & 1 \le \ell \le n-M \\ \sum_{k=\ell-(n-M)}^{M} \xi_{\ell-k} a_k, & n-M < \ell \le n. \end{cases} \tag{49}$$

As we proceed when we assumed that $B(L) = 1$, by standard inequalities, it suffices to show that

$$\frac{1}{n}\sum_{t=1}^{n} \left|\widehat{\varepsilon}_t - \mathring{\varepsilon}_t\right|^p = o_p(1) \tag{50}$$

$$\frac{1}{n}\sum_{t=1}^{n} \left|\mathring{\varepsilon}_t - \widetilde{\varepsilon}_t\right|^p + \frac{1}{n}\sum_{t=1}^{n} \left|\varepsilon_t - \widetilde{\varepsilon}_t\right|^p = o_p(1), \tag{51}$$

where $\widetilde{\varepsilon}_t = \sum_{\ell=0}^{t-1} \phi_\ell u_{t-\ell} + \sum_{\ell=t}^{n} \phi_\ell u_{n-(\ell-t)}$ and $\widehat{\varepsilon}_t$ given in (22).

Now, that the second term on the left of (51) is $o_p(1)$ follows proceeding as with the proof of (47). Next, the first term on the left of (51). Because $\mathring{\phi}_\ell - \phi_\ell = 0$ if $\ell \le M$, we have that that its expectation is bounded by

$$\begin{aligned} & \frac{1}{n}\sum_{t=M+1}^{n} E\left|\sum_{\ell=M+1}^{t-1}\left(\mathring{\phi}_\ell - \phi_\ell\right)u_{t-\ell}\right|^p + E\left|\sum_{\ell=t}^{n}\left(\mathring{\phi}_\ell - \phi_\ell\right)u_{n-(\ell-t)}\right|^p \\ & + \frac{1}{n}\sum_{t=1}^{M} E\left|\sum_{\ell=M+1}^{n}\left(\mathring{\phi}_\ell - \phi_\ell\right)u_{n-(\ell-t)}\right|^p \\ \le & \frac{1}{n}\sum_{t=M+1}^{n}\left\{\left(\sum_{\ell=M+1}^{t-1}\left|\mathring{\phi}_\ell - \phi_\ell\right|\right)^p + \left(\sum_{\ell=t}^{n}\left|\mathring{\phi}_\ell - \phi_\ell\right|\right)^p\right\} \\ & + \frac{1}{n}\sum_{t=1}^{M}\left(\sum_{\ell=M+1}^{n}\left|\mathring{\phi}_\ell - \phi_\ell\right|\right)^p \\ = & o(1) \end{aligned}$$

by (48) and that $E|u_t|^p < K$ and then because $\left|\mathring{\phi}_\ell - \phi_\ell\right|$ is a summable sequence by Lemma 7.

Finally (50), whose left term is bounded by

$$\begin{aligned} & \frac{K}{n}\sum_{t=1}^{M}\left\{\left|\sum_{\ell=0}^{t-1}\ddot{\phi}_\ell u_{t-\ell}\right|^p + \left|\sum_{\ell=t}^{M}\ddot{\phi}_\ell u_{n-(\ell-t)}\right|^p + \left|\sum_{\ell=M+1}^{n}\ddot{\phi}_\ell u_{n-(\ell-t)}\right|^p\right\} \\ & + \frac{K}{n}\sum_{t=M+1}^{n}\left\{\left|\sum_{\ell=0}^{M}\ddot{\phi}_\ell u_{t-\ell}\right|^p + \left|\sum_{\ell=M+1}^{t-1}\ddot{\phi}_\ell u_{t-\ell}\right|^p + \left|\sum_{\ell=t}^{n}\ddot{\phi}_\ell u_{n-(\ell-t)}\right|^p\right\}, \end{aligned}$$

where we have abbreviated $\widehat{\phi}_\ell - \mathring{\phi}_\ell =: \ddot{\phi}_\ell$. Now, that the fourth term is $o_p(1)$ follows because (59) yields that this term is

$$O_p\left(m^{-p/2}\right)\frac{K}{n}\sum_{t=M+1}^{n}\left|\sum_{\ell=0}^{M}|u_{t-\ell}|\right|^p = O_p\left(\left(\frac{M^2}{m}\right)^{p/2}\right) = o_p(1)$$

by Condition 3. Next by Lemma 6 part (ii), the sixth term is

$$O_p\left(m^{-p/2}\right)\frac{K}{n}\sum_{t=M+1}^{n}\left|\sum_{\ell=t}^{n}\left|\xi_{\ell-M}\right||u_{t-\ell}|\right|^p = O_p\left(m^{-p/2}\right)$$

because (48) and $E|u_t|^p < K$ imply that

$$E\sum_{t=M+1}^{n}\left|\sum_{\ell=t}^{n}\left|\xi_{\ell-M}\right||u_{t-\ell}|\right|^p \le K\sum_{t=M+1}^{n}\left\{\left(\sum_{\ell=t}^{n}(\ell-M)^{-1-(d_1\wedge d_2)}\right)^p\right\} = O(n)$$

because $(d_1\wedge d_2) > 0$. The fifth term follows by the same argument, whereas the first three terms are $o_p(1)$ proceeding as with the last three terms and that $M/n = o(1)$.

6.0.2. ***Proof of Proposition 2***.

Defining $\mathring{u}_t^* = \sum_{\ell=0}^{n+M}\mathring{\vartheta}_\ell\varepsilon_{t-\ell}^*$, where

$$\mathring{\vartheta}_\ell = \begin{cases}\sum_{k=0}^{\ell\wedge M} b_p\overline{\xi}_{\ell-p}, & 1\le\ell\le n\\ \sum_{k=\ell-n}^{M} b_p\overline{\xi}_{\ell-p}, & n<\ell\le n+M,\end{cases} \tag{52}$$

it suffices to show that

$$E^*\left|u_t^*-\mathring{u}_t^*\right|^p = o_p(1) \tag{53}$$

$$E^*\left|\mathring{u}_t^*-\widetilde{u}_t^*\right|^p = o_p(1). \tag{54}$$

We first examine (54). Because $\vartheta_\ell - \mathring{\vartheta}_\ell = 0$ if $\ell\le M$, we have that $\mathring{u}_t^* - \widetilde{u}_t^* = \sum_{\ell=M+1}^{n+M}\left(\mathring{\vartheta}_\ell-\vartheta_\ell\right)\varepsilon_{t-\ell}^*$, so that the left side of (54) becomes

$$E^*\left|\sum_{\ell=M+1}^{n+M}\left(\mathring{\vartheta}_\ell-\vartheta_\ell\right)\varepsilon_{t-\ell}^*\right|^p \le K\left(\sum_{\ell=M+1}^{n+M}\left|\mathring{\vartheta}_\ell-\vartheta_\ell\right|^2\right)^{p/2-1}\sum_{\ell=M+1}^{n+M}\left|\mathring{\vartheta}_\ell-\vartheta_\ell\right|^2E^*\left|\varepsilon_{t-\ell}^*\right|^p$$

by (28) because $\{\varepsilon_t^*\}_{t=1}^{n+M}$ is a random sequence with zero mean. Now, by definition we have that $\mathring{\vartheta}_\ell - \vartheta_\ell =: \overline{\xi}_\ell O\left(M^{-2}\right)$ by Lemma 9, so that we conclude that (54) holds true because $\left\{\overline{\xi}_\ell^2\right\}_{\ell\ge1}$ is a summable sequence and Corollary 1 implies that $E^*\left|\varepsilon_{t-\ell}^*\right|^p - E\left|\varepsilon_{t-\ell}\right|^p = o_p(1)$. So, to complete the proof it remains to show that (53) holds true. But this is the case because $u_t^* - \mathring{u}_t^* = \sum_{\ell=0}^{n+M}\left(\widehat{\vartheta}_\ell-\mathring{\vartheta}_\ell\right)\varepsilon_{t-\ell}^*$ and hence

$$E^*\left|u_t^*-\mathring{u}_t^*\right|^p \le K\left(\sum_{\ell=0}^{n+M}\left|\widehat{\vartheta}_\ell-\mathring{\vartheta}_\ell\right|^2\right)^{p/2-1}\sum_{\ell=0}^{n+M}\left|\widehat{\vartheta}_\ell-\mathring{\vartheta}_\ell\right|^2E^*\left|\varepsilon_{t-\ell}^*\right|^p$$

by (28). From here the proof follows by Lemma 8 since $\overline{\xi}_k^2$ is a summable sequence and Corollary 1.

6.0.3. ***Proof of Proposition 5***.

By the definition of u_t^* in (23), we obtain that

(55)
$$u_t^* = \sum_{\ell=0}^{n+M} \vartheta_\ell \varepsilon_{t-\ell}^* + \sum_{\ell=0}^{M} \left(\widehat{\vartheta}_\ell - \mathring{\vartheta}_\ell \right) \varepsilon_{t-\ell}^* + \sum_{\ell=M+1}^{n+M} \left(\widehat{\vartheta}_\ell - \mathring{\vartheta}_\ell \right) \varepsilon_{t-\ell}^* + \sum_{\ell=M+1}^{n+M} \left(\mathring{\vartheta}_\ell - \vartheta_\ell \right) \varepsilon_{t-\ell}^*,$$

where $\widehat{\vartheta}_\ell$ was given in (25) and because $\mathring{\vartheta}_\ell - \vartheta_\ell = 0$ if $\ell \leq M$. The last term on the right of (55) is $o_{p^*}(1)$, since its second moment is

$$\sum_{\ell=M+1}^{n+M} \left(\mathring{\vartheta}_\ell - \vartheta_\ell \right)^2 E^* \varepsilon_{t-\ell}^{*2} = o_p(1)$$

by summability of $\left\{ \left| \mathring{\vartheta}_\ell - \vartheta_\ell \right|^2 \right\}_{\ell \geq 1}$ by Lemma 9 and Corollary 1 implies that $E^* \varepsilon_{t-\ell}^{*2} - E\varepsilon_{t-\ell}^2 = o_p(1)$. Next, the second and third terms on the right of (55) are also $o_{p^*}(1)$ as we now show. Indeed, the second moment is

$$\sum_{\ell=0}^{M} \left(\widehat{\vartheta}_\ell - \mathring{\vartheta}_\ell \right)^2 E^* \varepsilon_{t-\ell}^{*2} + \sum_{\ell=M+1}^{n+M} \left(\widehat{\vartheta}_\ell - \mathring{\vartheta}_\ell \right)^2 E^* \varepsilon_{t-\ell}^{*2}.$$

From here we conclude by Lemma 8 parts (i) and (ii) respectively and again by Corollary 1. Thus, Markov inequality implies that the last three terms on the right of (55) are $o_{p^*}(1)$, so that

$$u_t^* = \sum_{\ell=0}^{n+M} \vartheta_\ell \varepsilon_{t-\ell}^* + o_{p^*}(1).$$

Now, Proposition 4 and Cramér-Wold's device imply that for any $u \in \mathbb{R}$,

$$P^* \left(\sum_{\ell=0}^{n+M} \vartheta_\ell \varepsilon_{t-\ell}^* \leq u \right) = P \left(\sum_{\ell=0}^{n+M} \vartheta_\ell \varepsilon_{t-\ell} \leq u \right) + o_p(1).$$

Since $\left\{ \vartheta_\ell^2 \right\}_{\ell \geq 1}$ is summable, $E\left(\sum_{\ell=n+M+1}^{\infty} \vartheta_\ell \varepsilon_{t-\ell} \right)^2 = o(1)$ and Markov's inequality yields that

$$u_t = \sum_{\ell=0}^{\infty} \vartheta_\ell \varepsilon_{t-\ell} = \sum_{\ell=0}^{n+M} \vartheta_\ell \varepsilon_{t-\ell} + o_p(1).$$

Gathering the last three displayed expressions we conclude the proof of the proposition.

7. LEMMAS.

LEMMA 1. *For any $H \geq 2$ and under Condition 1, we have that if $d_1 > 0$*

$$\textbf{(a)} \quad \frac{\widehat{\pi}_k - \pi_k}{\pi_k} = \left(\sum_{h=1}^{H-1} \left(\widehat{d}_1 - d_1 \right)^h g_k^h(d_1) \right) + O_p \left(\left| \widehat{d}_1 - d_1 \right|^H \right) \log^H k$$

$$\textbf{(b)} \quad \frac{\check{\pi}_k - \overline{\pi}_k}{\overline{\pi}_k} = \left(\sum_{h=1}^{H-1} \left(\widehat{d}_1 - d_1 \right)^h g_k^h(d_1) \right) + O_p \left(\left| \widehat{d}_1 - d_1 \right|^H \right) \log^H k,$$

where $g_k(d_1) = d_1^{-1} + \sum_{\ell=2}^{k} \ell^{-1}$, *whereas if* $d_1 = 0$

$$(\mathbf{c}) \quad \widehat{\pi}_k = \frac{\log k}{k}\widehat{d}_1 + \frac{1}{k}O_p\left(\left|\widehat{d}_1\right|^2\right) + O_p\left(\left|\widehat{d}_1\right|^H\right)\frac{\log^H k}{k}$$

and the $O_p(\circ)$ *is uniformly in* k.

We shall denote proportional by "$\propto$" that is "$c_n \propto \ell_n$" means that $K^{-1}\ell_n \leq c_n \leq K\ell_n$ for some finite and positive constant K.

LEMMA 2. *For any* $H \geq 2$ *and under Condition* 1, *we have that if* $|d_2| > 0$

$$(\mathbf{a})\ \widehat{\tau}_k - \tau_k = \tau_k\left(\sum_{h=1}^{H-1}\left(\widehat{d}_2 - d_2\right)^h g_k^h(d_2)\right) + O_p\left(\left|\widehat{d}_2 - d_2\right|^H \frac{\log^H k}{k^{1-d_2}}\right),$$

where $g_k(d_2) = d_2^{-1} + \sum_{\ell=2}^{k} \ell^{-1}$, *whereas when* $d_2 = 0$,

$$(\mathbf{b})\ \widehat{\tau}_k = \left(\frac{1}{k}\sum_{h=1}^{H-1}\widehat{d}_2^h g_k^h(0)\right) + O_p\left(\left|\widehat{d}_2\right|^H\right)\frac{\log^H k}{k},$$

where $g_k(0) \propto (1 + \log k)$.

LEMMA 3. *Under Condition* 1, *we have that for any integer* $H \geq 2$

$$(a)\ \widehat{\xi}_k - \xi_k = \xi_k\sum_{h=1}^{H-1}\left\|\widehat{d} - d\right\|^h g_k^h(d) + O_p\left(\left\|\widehat{d} - d\right\|^H \frac{\log^H k}{k^{1+(d_1\wedge d_2)}}\right);\ d_1, d_2 > 0$$

$$(b)\ \check{\xi}_k - \bar{\xi}_k = \bar{\xi}_k\sum_{h=1}^{H-1}\left\|\widehat{d} - d\right\|^h g_k^h(d) + O_p\left(\left\|\widehat{d} - d\right\|^H \frac{\log^H k}{k^{1+(d_1+d_2)}}\right);\ d_1, d_2 < 0,$$

where $d = (d_1, d_2)$, $g_k(\|d\|) \propto \left(\|d\|^{-1} + \log k\right)$, $\|a\|$ *denotes the norm of the vector* a *and* $O_p(\circ)$ *is uniformly in* k.

For the next lemma we shall denote by $\{\delta_\ell\}_{\ell=1}^M$ and $\left\{\widehat{\delta}_\ell\right\}_{\ell=1}^M$ either $\{b_\ell\}_{\ell=1}^M$ and $\left\{\widehat{b}_\ell\right\}_{\ell=1}^M$ or $\{a_\ell\}_{\ell=1}^M$ and $\{\widehat{a}_\ell\}_{\ell=1}^M$.

LEMMA 4. *Under Conditions* 1, 2 *and* 4, *we have that, uniformly in* $\ell = 1, ..., M$,

$$\widehat{\delta}_\ell - \delta_\ell = \sum_{p=1}^{M}\kappa_{n,p}\upsilon_{n,\ell,p} + \frac{1}{M}\sum_{p=1}^{M}|\kappa_{n,p}\upsilon_{n,\ell,p}| + O_p\left(m^{-3/2}\right),$$

where $|\upsilon_{n,\ell,p}| < K$, $\{\kappa_{n,p}\}_{p=1}^M$ *is a triangular array sequence of random variables such that* $E\left|\sum_{p=1}^{M}\kappa_{n,p}\upsilon_{n,\ell,p}\right|^r = O\left(n^{-r/2}\right)$ *for any* $r \geq 2$ *such that* $E|\varepsilon_t|^{2r} < \infty$, $E\left(\sum_{p=1}^{M}\kappa_{n,p}\upsilon_{n,\ell_1,p}\sum_{p=1}^{M}\kappa_{n,p}\upsilon_{n,\ell_2,p}\right) = O\left(n^{-1}\delta_{|\ell_2-\ell_1|}\right)$ *and* $E\sum_{p=1}^{M}|\kappa_{n,p}\upsilon_{n,\ell,p}| = O\left(m^{-1/2}\right)$.

REMARK 6. *Notice that one consequence of Lemma 4 is that*

$$\sup_{\ell=1,\dots,M}\left|\widehat{\delta}_\ell - \delta_\ell\right| = O_p\left(m^{-1/2}\right). \tag{56}$$

LEMMA 5. *For* $0 \le r \le n$,

$$E^*\left(u_t^* u_{t+r}^*\right) = \widehat{\sigma}_\varepsilon^2 \sum_{k=0}^{n+M-r} \widehat{\vartheta}_k \widehat{\vartheta}_{k+r}. \tag{57}$$

First by definitions of $\widehat{\phi}_k$ and $\mathring{\phi}_k$ in (20) and (49) respectively, we have that

$$\widehat{\phi}_k - \mathring{\phi}_k = \sum_{\ell=1\vee k-(n-M)}^{k\wedge M} \widehat{\xi}_{k-\ell}\left(\widehat{a}_\ell - a_\ell\right) + \sum_{\ell=1\vee k-(n-M)}^{k\wedge M} \left(\widehat{\xi}_{k-\ell} - \xi_{k-\ell}\right) a_\ell, \qquad k \le n. \tag{58}$$

LEMMA 6. *Under Conditions* 1 *and* 3, *we have that*

$$\text{(i)} \quad \widehat{\phi}_k - \mathring{\phi}_k = \widetilde{\varsigma}_{n,M} + \varsigma_{n,M}\xi_k, \qquad k \le M$$

$$\text{(ii)} \quad \widehat{\phi}_k - \mathring{\phi}_k = \varsigma_{n,M}\xi_{k-M}, \qquad M < k < n,$$

where $\varsigma_{n,M} = O_p\left(m^{-1/2}\right)$ *and* $\varsigma_{n,M} = O_p\left(n^{-1/2}\right)$ *independent of* k.

REMARK 7. *A consequence of Lemma 6 together with* (56) *is that*

$$\sup_{k=1,\ldots,n}\left|\widehat{\phi}_k - \mathring{\phi}_k\right| = O_p\left(m^{-1/2}\right). \tag{59}$$

The next lemma examines the behaviour of

$$\begin{aligned} \mathring{\phi}_k - \phi_k &= \sum_{\ell=M+1}^{k} \xi_{k-\ell} a_\ell & M < k \le n-M \\ &= -\sum_{\ell=0}^{k-(n-M)} \xi_{k-\ell} a_\ell - \sum_{\ell=M+1}^{k} \xi_{k-\ell} a_\ell & n-M < k \le n \end{aligned}$$

LEMMA 7. *Under Condition* 1, *we have that*

$$\text{(i)} \quad \mathring{\phi}_k - \phi_k = O\left(M^{-3/2}\right)\xi_k, \qquad M < k \le n-M$$

$$\text{(ii)} \quad \mathring{\phi}_k - \phi_k = O(1)\,\xi_k, \qquad n-M < k < n.$$

Next by definition of $\widehat{\vartheta}_k$ and $\mathring{\vartheta}_k$ in (25) and (52) respectively, we have that

$$\widehat{\vartheta}_k - \mathring{\vartheta}_k = \sum_{\ell=1}^{k\wedge M} \check{\xi}_{k-\ell}\left(\widehat{b}_\ell - b_\ell\right) + \sum_{\ell=1}^{k\wedge M}\left(\check{\xi}_{k-\ell} - \overline{\xi}_{k-\ell}\right) b_\ell, \qquad 1 \le k \le n+M \tag{60}$$

LEMMA 8. *Under Conditions* 1 *and* 3, *we have that*

$$\text{(i)} \quad \widehat{\vartheta}_k - \mathring{\vartheta}_k = \sum_{\ell=0}^{k\wedge M} \overline{\xi}_{k-\ell}\left(\widehat{b}_\ell - b_\ell\right) + \left(\varsigma_n + o_p\left(n^{-1}\right)\right)\overline{\xi}_k \qquad 0 \le k \le M$$

$$\text{(ii)} \quad \widehat{\vartheta}_k - \mathring{\vartheta}_k = \sum_{\ell=0}^{k\wedge M} \overline{\xi}_{k-\ell}\left(\widehat{b}_\ell - b_\ell\right) + \left(\varsigma_n + o_p\left(n^{-1}\right)\right)\overline{\xi}_{k-M} \qquad M < k \le n+M,$$

where $\varsigma_n = O_p\left(m^{-1/2}\right)$ *independent of* k.

REMARK 8. *A consequence of the previous lemma together with (56) is that*

$$\sup_{k=1,\ldots,n}\left|\widehat{\vartheta}_k-\mathring{\vartheta}_k\right|=O_p\left(m^{-1/2}\right). \tag{61}$$

By definition of ϑ_k and $\mathring{\vartheta}_k$ in (2) and (52) respectively, we have that

$$\begin{aligned}\mathring{\vartheta}_k-\vartheta_k &= \sum_{\ell=M+1}^{k}\overline{\xi}_{k-\ell}b_\ell && M<k\le n\\ &= -\sum_{\ell=0}^{k-n}\overline{\xi}_{k-\ell}b_\ell-\sum_{\ell=M+1}^{k}\overline{\xi}_{k-\ell}b_\ell && n<k\le n+M\end{aligned}$$

LEMMA 9. *Under Condition 1, we have that*

(i) $\mathring{\vartheta}_k-\vartheta_k=O\left(M^{-2}\right)\overline{\xi}_k$, $\quad M<k\le n$

(ii) $\mathring{\vartheta}_k-\vartheta_k=O\left(1\right)\overline{\xi}_k$, $\quad n<k<n+M.$

REFERENCES

[1] ARTECHE, J.M. (2000): "Gaussian semiparametric estimation in seasonal/cyclical long memory time series," *Kybernetika*, **36**, 279-310.

[2] BELTRÃO, K.I. AND BLOOMFIELD, P. (1987), "Determining the bandwidth of a kernel spectrum estimate," *Journal of Time Series Analysis*, **8**, 21–38.

[3] BERK, K.N. (1974): "Consistent autoregression spectral estimates," *Annals of Statistics*, **2**, 489-502.

[4] BICKEL, P.J. AND FREEDMAN, D.A. (1981): "Some asymptotic theory for the bootstrap," *Annals of Statistics*, **6**, 1196–1217.

[5] BROCKWELL, P.J. AND DAVIS, R.A. (1991): "Time Series: Theory and Methods," Springer, New York.

[6] BÜHLMANN, P. (1997): "Sieve-bootstrap for time series," *Bernoulli*, **3**, 123-148.

[7] BÜHLMANN, P. (2002): "Bootstraps for time series," *Statistical Science*, **17**, 52-72.

[8] DAHLHAUS, R. AND JANAS, D. (1996): "A frequency domain bootstrap for ratio statistics in time series analysis," *Annals of Statistics*, **24**, 1934-1963.

[9] DEHLING, H. AND TAQQU, M.B. (1989): "The empirical process of some long-memory dependent sequences with an application to U-statistics," *Annals of Statistics*, **17**, 1767–1783.

[10] ELAYDI, S. (2005): "An introduction to difference equations," 3rd Edition. Springer Verlag.

[11] EFRON, B. (1979): "Bootstrap methods: Another look at the jackknife," *Annals of Statistics*, **7**, 1-26.

[12] FRANKE, J. AND HÄRDLE, W. (1992): "On bootstrapping kernel spectral estimates," *Annals of Statistics*, **20**, 121-145.

[13] GIACOMINI, R., POLITIS, D.N. AND WHITE, H. (2013): "A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators," *Econometric Theory*, **29**, 567-589.

[14] GIRAITIS, L. AND SURGAILIS, D. (1994): "A central limit theorem for the empirical process of a long memory linear sequence," Mimeo. Universität Heidelberg. Beltläge zur Statistik, nr. 24.

[15] HALL, P. (1992): "The bootstrap and Edgeworth expansion," Springer, New York.

[16] HIDALGO, J. (2005): "Semiparametric estimation for stationary processes whose spectra have an unknown pole," *Annals of Statistics*, **33**, 1843-1889.

[17] HIDALGO, J. (2003): "An alternative bootstrap to moving blocks for time series regression models," *Journal of Econometrics*, **117**, 369–399.

[18] HIDALGO, J. (2008): "Specification testing for regression models with dependent data," *Journal of Econometrics*, **143**, 143-165.

[19] HIDALGO, J. (2019): "Model diagnostics in time series regression models," Mimeo.

[20] HIDALGO, J. AND ROBINSON, P.M. (2002): "Adapting for unknown autocorrelation in the disturbance ", *Econometrica*, **70**, 1545-1581.

[21] HIDALGO, J. AND YAJIMA, Y. (2002): "Prediction in the frequency domain under long-range processes with application to the signal extraction problem," *Econometric Theory*, **18**, 584-624.

[22] HOROWITZ, J.L. (1997): "Bootstrap methods in econometrics," in *Advances in Economics and Econometrics: Theory and Applications, Seventh World Congress*, (ed. by D.M. Kreps and K.R. Wallis), Cambridge: Cambridge University Press, 188-222.

[23] HURVICH, C.M. AND ZEGER, S.J. (1987): "Frequency domain bootstrap methods for time series," New York University Working paper.

[24] KIRSCH, C. AND POLITIS, D.N. (2011): "TFT-Bootstrap: Resampling time series in the frequency domain to obtain replicates in the time domain," *Annals of Statistics*, **39**, 1427–1470.

[25] KOUL, H. L., BAILLIE, R. AND SURGAILIS, D. (2004): "Regression model fitting with a long memory covariate process," *Econometric Theory*, **20**, 485–512.

[26] KREISS, J.P. (1988): "Asymptotical inference for a class of stochastic processes," Habilitationsschrift, Univ. Hamburg.

[27] KREISS, J.P. AND PAPARODITIS, E. AND POLITIS, D.N. (2011): "On the range of validity of the autoregressive sieve bootstrap," *Annals of Statistics*, **39**, 2103-2130.

[28] KÜNSCH, H.R. (1989): "The jackknife and the bootstrap for general stationary observations," *Annals of Statistics*, **17**, 2103-2130.

[29] LAHIRI, S.N. (2003): "Resampling methods for dependent data," Springer, New York.

[30] LOBATO, I. AND ROBINSON P.M. (1997): "A nonparametric test for $I(0)$," *Review of Economic Studies,* **65***,* 475-495.

[31] MARINUCCI, D. AND ROBINSON, P.M. (1999): "Alternative forms of fractional Brownian motion," *Journal of Statistical Planning and Inference,* **80**, 111-122.

[32] MARINUCCI, D. AND ROBINSON, P.M. (2000): "Weak convergence of multivariate fractional processes," *Stochastic Processes and their Applications*, **86**, 103-120.

[33] POLITIS, D.N. (2003): "The impact of bootstrap methods on time series analysis," *Statistical Science*, **18**, 219-230.

[34] PRESS H. AND TUKEY, J.W. (1956): "Power spectral methods of analysis and their application to problems in airplane dynamics," *Bell Telephone System Monograph* **2606**.

[35] ROBINSON, P.M. (1991): "Automatic frequency domain inference on semiparametric and nonparametric models," *Econometrica*, **59**, 1329-1363.

[36] ROBINSON, P.M. (1995a): "Log-periodogram regression for time series with long range dependence," *Annals of Statistics*, **23**, 1048-1072.

[37] ROBINSON, P.M. (1995b): "Gaussian semiparametric estimation of long-range dependence," *Annals of Statistics*, **23**, 1630-1661.

[38] ROBINSON, P. M. AND HIDALGO, J. (1997): "Time series regressions with long-range dependence," *Annals of Statistics*, **25**, 77-104.

[39] SHAO, J. AND TU, D. (1995): "The jackknife and bootstrap." Springer Verlag. Berlin.

[40] STUTE, W. (1997): "Nonparametric model checks for regression," *Annals of Statistics*, **25**, 613–641.

[41] WU, W.B. (2003): "Empirical processes of long-memory sequences," *Bernoulli*, **9**, 809–831.

[42] YONG, C.H. (1974): "Asymptotic behaviour of trigonometric series," Chinese University of Hong Kong, Hong Kong.