Bootstrap log-average sample spectral estimator of the fractional parameter in ARFIMA models

Valderio A. Reisen\textsuperscript{a,*}, Glaura C. Franco \textsuperscript{b} and Giovanni Comarela \textsuperscript{a}

\textsuperscript{a}Department of Statistics, UFES, ES, Brazil.
\textsuperscript{b}Department of Statistics, UFMG, MG, Brazil

October 2, 2008

Abstract

A bootstrap procedure is employed to obtain bootstrap replications of the periodogram in order to calculate a bootstrap log average sample spectral estimator of the fractional parameter $d$ in ARFIMA($p,d,q$) models. This procedure follows a semiparametric regression approach and a Monte Carlo study is carried out to compare the finite sample properties of the proposed estimator with the classical semiparametric regression approach based on the periodogram function. The proposed estimator presented the smallest mean square error and approximate the real value of the parameter when the size of the series increases.

Keywords: ARFIMA models, local bootstrap, average periodogram.

1 Introduction

Many estimators for the fractional parameter $d$ in long memory time series have already been proposed in the literature. Among them are the semiparametric procedures, which include a wide variety of estimators based on the ordinary least squared (OLS) method. These procedures require the use of the spectral density parameterized within a neighborhood of zero frequency. Some references on this subject include the works of Geweke and Porter-Hudak (1983), Reisen (1994) and Robinson (1995), among others. An overview of long-range dependence processes can be found in Beran (1994) and Doukham et al. (2003).

*Corresponding author: Valderio A. Reisen. Address: Departamento de Estatística - CCE - UFES - Av. Fernando Ferrari, s/n - Vitória - ES - Brazil - CEP: 29065-900. E-mail: valderio@cce.ufes.br
An asymptotically unbiased, but inconsistent, estimate of the spectral density is given by the periodogram function (Priestley, 1981). To overcome the problem of consistency, Reisen et al. (2008) have proposed an averaged periodogram estimator, based on the idea of the averaging periodogram, computed when the series is divided into blocks, calculating the periodogram for each block and averaging these to obtain the average periodogram spectral estimator. See for example, Priestley (1981), Rao (1993) and Hernandez-Flores et al. (1999) for the properties of the averaging periodogram.

In their work, Reisen et al. (2008) use the averaging periodogram spectral estimator to obtain an estimator of the memory parameter, \( d \), of the ARFIMA process, following the procedure of Geweke and Porter-Hudak (1983), where the spectral function is replaced by the averaging periodogram in the regression equation of the logarithm of the spectral density. They prove some asymptotic properties of this estimator and perform empirical investigation to give evidence of its good performance.

In this work another way of obtaining averaging periodograms is proposed, using the bootstrap technique. There are many different approaches to perform the bootstrap in time series and particularly in the long-memory domain. The works of Grau-Carles (2004), de Peretti (2002), Franco and Reisen (2004, 2007) and Arteche and Orbe (2005) are only some of the references that can be cited in the long-range time series models. The idea here is to get replicated periodograms of the original series, using the bootstrap, and calculate the average of these bootstrap periodograms. Here, the local bootstrap introduced by Shi (1991) is used to this purpose. Other authors have also used the local bootstrap applied to periodogram statistics, such as Paparoditis and Politis (1999), and to ARFIMA processes (Silva et al. (2006) and Arteche and Orbe (2008)).

The paper is organized as follows: In Section 2 the models and some asymptotic properties of the average periodogram are presented. The theoretical issues discussed in this section are the background theory for the local-average bootstrap periodogram estimator introduced in Section 3 which deals with the local and the average local bootstrap approaches. Section 4 introduces the proposed memory parameter estimator based on the average local bootstrap. Section 5 shows some simulation results and Section 6 concludes the work.

## 2 The model and properties

Let \( \{X_t\} \) be the general linear process given by

\[
X_t = \psi(B)\epsilon_t = \sum_{j=-\infty}^{\infty} \psi_j B^j \epsilon_t, \quad j, t \in \mathbb{Z}
\]  

(1)
where $B$ is the backshift operator ($B^j\epsilon_t = \epsilon_{t-j}$) and \(\{X_t\}\) satisfies the assumptions:

**A1** \(\sum_{j=-\infty}^{\infty} |j|^\alpha |\psi_j| < \infty\), $\psi_0 = 1$ and $\alpha > 0$;

**A2** \(\{\epsilon_t\}\) is a random process with zero mean, constant variance $\sigma^2_\epsilon$ and $E(\epsilon^4_t) < \infty$.

These assumptions guarantee that \(\{X_t\}\) is a covariance stationary process, i.e., $Cov(X_t, X_{t+j}) = \gamma(j)$ do not depend on $t$ and the process has absolutely summable autocovariances ($\sum_{j=0}^{\infty} |\gamma(j)| < \infty$). They are also required for the asymptotic properties of the spectral density estimator discussed in the next pages.

**A3** The spectral density of \(\{X_t\}\), $f_X(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma(j) e^{-ij\omega} = \frac{\sigma^2_\epsilon}{2\pi} |\psi(e^{-i\omega})|^2$, is nonvanishing and continuously differentiable with bounded derivative for $-\pi \leq \omega \leq \pi$.

The stationary Autoregressive Moving-Average (ARMA) process is an example of the above process since its autocovariance function satisfies $|\gamma_{ARMA}(j)| \leq C\tau^{|j|}$ with constants $C$ and $\tau$ ($0 < \tau < 1$). The ARMA process is usually defined as short-memory process.

Suppose $x_1, ..., x_n$ is a partial realization of \(\{X_t\}\). Hence, the periodogram function is defined as $I_x(\omega) = (2\pi n)^{-1} \sum_{t=1}^{n} |x_t e^{i\omega t}|^2$. It follows that

\[
I_x(\omega) = 2\pi f_X(\omega) \frac{I_x(\omega)}{\sigma_\epsilon^2} + H(\omega) \tag{2}
\]

where $E[|H(\omega)|^2] = O\left(\frac{1}{n\pi}\right)$ ($\xi > 0$) is uniform in $\omega \in [-\pi, \pi]$ (Theorem 6.2.2 in Priestley (1981)). From equation (2) and Theorem 6.1.1 in Priestley (1981), asymptotic sample properties of $\frac{I_x(\omega)}{f_X(\omega_j)}$ are derived and they are summarized as follows. If \(\{\epsilon_t\}\) are normally distributed, for a fixed set of values of the Fourier frequencies $\omega_j = \frac{2\pi j}{n}, j = 1, ..., [n/2]$, where $[\cdot]$ means the integer part, $0 < \omega_j < \pi$, asymptotically the set of variables $\frac{I_x(\omega)}{f_X(\omega_j)}$ are independently distributed, each distributed as $\chi^2_{\frac{2}{\pi}}$. At $\omega = 0$ and $\pi$ the distributions are $\chi^2_1$ (for details see Priestley (1981) and Rao (1993)). Since $E[\chi^2_2] = 2$ and $\text{Var}[\chi^2_2] = 4$ then $E\left[\frac{I_x(\omega)}{f_X(\omega_j)}\right] \rightarrow 1$ and $\text{Var}[\frac{I_x(\omega)}{f_X(\omega_j)}]\rightarrow (1 + \delta(\omega))$ as $n \rightarrow \infty$, where $\delta(\omega) = 1$ if $\omega = 0$ (mod $\pi$) and 0 otherwise. This establishes the unbiased and inconsistency properties of $I_x(\omega_j)$.

Suppose now that \(\{X_{t,l}\}, l = 1, ..., r, r \in \mathbb{Z}\), are $r$ independent realizations of the general linear process \(\{X_t\}\), with $\gamma_l(j) = \gamma(j)$ for all $j$. Let $x_{1,1}, ..., x_{1,n}$ be a partial replication of \(\{X_{t,l}\}\). For a fixed $\omega_j$, the average periodogram is defined by

\[
\bar{I}(\omega_j) = \frac{1}{r} \sum_{l=1}^{r} I_x(\omega_j) \tag{3}
\]

where $I_x(\omega_j)$ is the periodogram of the sample \(\{x_{1,l}\}\). Then, it follows immediately from the asymptotic periodogram properties that, as $n \rightarrow \infty$, $\frac{I_x(\omega_j)}{f_X(\omega_j)} \sim \frac{\chi^2_{\frac{2}{2\pi}}}{2\pi}$ and
\( \bar{I}(\omega_j) \) and \( \bar{I}(\omega_k) \) are independent for \( j \neq k \). From these, the following proposition is established.

**Proposition 1.** Let \( \bar{I}(\omega_j) \) be, at the Fourier frequency \( \omega_j \in (0, \pi) \), the average of the periodogram of \( r \) independent partial realizations \( \{x_{l,t}\}, t = 1, ..., n, l = 1, ..., r, \) of the process \( \{X_t\} \) defined in (1), with spectral density \( f_X(\omega_j) \) and \( \{\epsilon_t\} \) a Gaussian white noise process with zero mean and constant variance \( \sigma^2 \). For \( r \to \infty \) and \( n \to \infty \):

i. \( \mathbb{E} \left[ \frac{\bar{I}(\omega_j)}{f_X(\omega_j)} \right] \to 1; \)

ii. \( r \text{Var} \left[ \frac{\bar{I}(\omega_j)}{f_X(\omega_j)} \right] \to 1 + \delta(\omega); \)

iii. \( \sqrt{r} \frac{\bar{I}(\omega_j) - \mathbb{E}[\bar{I}(\omega_j)]}{f_X(\omega_j)} \xrightarrow{d} N(0,1). \)

Proofs: (i) and (ii) are straightforward obtained from the asymptotic spectral theory and (iii) is directly proved from the central limit theorem.

Non-linear functions of variables are usually required in many applications. The logarithm is one of the most useful non-linear transformations because, besides stabilizing the variance, it has also the property of approximating the data to normality. In the case of the average of periodogram, this spectral estimator is always positive and \( \mathbb{E}[\bar{I}(\omega_j)] \) and \( \text{Var}[\bar{I}(\omega_j)] \) are, respectively, proportional to \( f_X(\omega_j) \) and \( f_X(\omega_j)^2 \). Thus a logarithmic transformation is clearly advised. Here, the major interest in the log transformation is due to the model parameter estimation discussed in the next pages.

Since \( \bar{I}(\omega_j) \), for each \( \omega_j \in (0, \pi) \), is distributed as \( \frac{f_X(\omega_j)}{2r} \chi^2_2 \), then the expectation and variance of log \( \frac{\bar{I}(\omega_j)}{f_X(\omega_j)} \) are, respectively, \( \mathbb{E} \left[ \log \frac{\bar{I}(\omega_j)}{f_X(\omega_j)} \right] = \varphi(r) - \log(r) \) and \( \zeta = \text{Var} \left[ \log \frac{\bar{I}(\omega_j)}{f_X(\omega_j)} \right] = \varphi'(r) \) where \( \varphi(\cdot) \) is the digamma function, \( \varphi(u) = \frac{\partial}{\partial u} \log \Gamma(u) \), and \( \Gamma(.) \) is the gamma function. In the case where \( r = 1 \), \( \varphi(1) = -0.577216 \) (Euler’s constant) and \( \zeta = \pi^2/6 \). At \( r = 5, 10 \) and 100, the variance, with five decimal places, is 0.22132, 0.10516, 0.01005, respectively. This indicates that \( \zeta \) decreases with \( r \).

Theorem 1 in Hurvich et al. (2001) states that for any fixed \( \zeta \), \( \lim_{r \to \infty} (r + \zeta) \varphi'(r) = \frac{\Gamma(4\zeta+1)\Gamma^3(\zeta+1)}{\Gamma^4(2\zeta+1)} \) (see also Moulines and Soulier (2003)). Hence, for \( \zeta = 0 \) and large \( r \), \( \zeta = O(r^{-1}) \).

The theoretical issues discussed in this section for the average periodogram are the background theory for the local-average bootstrap periodogram estimator introduced in the next section.
3 Local-average bootstrap for periodogram statistics

One way of obtaining replications of the process is performing the bootstrap (Efron, 1979). A bootstrap procedure for the periodogram of a process satisfying (1) was proposed by Paparoditis and Politis (1999). The resampling scheme proposed by the authors relies on the asymptotic independence of the periodogram ordinates and they claim that the generated bootstrap periodogram ordinates are asymptotically independent, that is, the weak dependence between the periodogram ordinates does not affect the asymptotic distribution of their proposed bootstrap periodogram function. Some of their main results are in what follows and they will be the base for the local-average bootstrap procedure here proposed to obtain an estimate of the spectral density of a weak stationary process that satisfies the model specified by (1).

Let \( \{X_t\} \) be a process with MA(\( \infty \)) representation as given in (1) that satisfies \( A_1 - A_3 \).

Given the partial realization \( x_1, ..., x_n \) of \( \{X_t\} \), the local bootstrap algorithm generates bootstrap periodogram replicates \( I^*(\omega_j), j = 1, ..., T \), where \( T = \lceil n/2 \rceil \), of the periodogram as follows.

1 Select a resampling width \( \kappa_n = \kappa(n) \in \mathbb{N} \) and \( \kappa_n \leq \lfloor T/2 \rfloor \).

2 Define i.i.d. discrete random variables \( j_1, ..., j_T \) taking values in the set \( \{-\kappa_n, -\kappa_n + 1, ..., \kappa_n\} \) with probability \( P(j_i = s) = p_{\kappa_n,s} \), for \( s = 0, \pm 1, ..., \pm \kappa(n) \) such that \( p_{\kappa_n,s} = p_{\kappa_n,-s} \).

3 The bootstrap periodogram is then defined by \( I^*(\omega_j) = I_X(\omega_{j_i+j}) \) for \( j = 1, ..., T \), \( I^*(\omega_j) = I^*(-\omega_j) \), for \( \omega_j < 0 \) and \( I^*(0) = 0 \).

The simplest case of \( p_{\kappa_n,s} \) is \( p_{\kappa_n,s} = \frac{1}{2\kappa_n+1} \) which uniform probability is assigned to each periodogram ordinate in \( s \). The periodogram estimator obtained from the above procedure maintains the same asymptotic properties as the periodogram from the observed series. To reach the asymptotic distribution of \( I^*(\omega_j) \), Paparoditis and Politis (1999) also stated the following assumption.

\textbf{A4} \( \kappa_n \to \infty \) as \( n \to \infty \) but such that \( \kappa_n = o(n) \) and the sequence \( \{p_{\kappa_n,s} : -\kappa_n \leq s \leq \kappa_n\} \) satisfies \( \sum_{s=-\kappa_n}^{\kappa_n} p_{\kappa_n,s} = 1 \), \( p_{\kappa_n,s} = p_{\kappa_n,-s} \) and \( \sum_{s=-\kappa_n}^{\kappa_n} p_{\kappa_n,s}^2 \to 0 \), as \( \kappa_n \to \infty \).

\textbf{Proposition 2.} Assume (A1)-(A4). Then as \( n \to \infty \)

\textbf{2.a} \( \mathbb{E}[I^*(\omega)/f(\omega)] \xrightarrow{p} 1; \)

\textbf{2.b} \( \text{Var}[I^*(\omega)/f(\omega)] \xrightarrow{p} 1 + \delta(\omega), \) where \( \delta(\omega) = 1 \) if \( \omega = 0 \ (mod \ \pi) \) or 0 otherwise;
Note: In the above proposition, \( E^*(\cdot) \) and \( \text{Var}^*(\cdot) \) refer to, respectively, the expectation and variance of the bootstrap periodogram \( \hat{I}^*(\omega) \) calculated with respect to the bootstrap probability \( p_{n,s} \). The proof of Proposition 2 is straightforward obtained by using the results on pages 196-197 in Paparoditis and Politis (1999) which are based on the fact that the bootstrap-local periodogram can be thought of as Kernel estimator of the spectral density.

**Corollary 1.** Assume (A1)-(A4). Then as \( n \to \infty \), \( \frac{\hat{I}^*(\omega)}{\hat{I}(\omega)} \) has the same asymptotic distribution of \( \frac{I(\omega)}{f(\omega)} \), i.e, \( \frac{\hat{I}^*(\omega)}{\hat{I}(\omega)} \overset{d}{\to} \chi_2^2 \nu \), where \( \nu = 2 - \delta(\omega) \) and \( \delta(\omega) \) is defined as previously.

The above corollary is a consequence of Theorem 2.1 presented in Paparoditis and Politis (1999).

Let

\[
\bar{I}^*(\omega_j) = \frac{\sum_{i=1}^{r} I^*_i(\omega_j)}{r}
\]

be the average of \( r \) bootstrap replications \( I^*_i(\omega_j) \), \( i = 1, \ldots, r \), of \( I_x(\omega_j) \) where \( \{x_t\}_{t=1,...,n} \) is a data set generated from the process \( \{X_t\} \) satisfying conditions of model (1).

In consequence of the previous statements, the following proposition is established for the local-average bootstrap periodogram estimator.

**Proposition 3.** Assume (A1)-(A4). Let \( r^* \ (r^* \leq r) \) be a number of different bootstrap periodogram replicates such that \( r^* \leq 2\kappa_n + 1 \). Then for a large \( r^* \) and as \( n \to \infty \)

(i) \( E^*[\frac{\bar{I}^*(\omega_j)}{f_X(\omega_j)}] \to 1; \)

(ii) \( r^* \text{Var}^*[\frac{\bar{I}^*(\omega_j)}{f_X(\omega_j)}] \to \{\delta(\omega) + 1\}. \)

(iii) \( \frac{\bar{I}^*(\omega)}{f(\omega)} \) has the same asymptotic distribution of \( \frac{I(\omega)}{f(\omega)} \), i.e,

\[
\sqrt{r^*} \left[ \frac{\bar{I}^*(\omega_j)}{f_X(\omega_j)} - E[\bar{I}^*(\omega_j)] \right] \overset{d}{\to} N(0,1).
\]

Note: From the above, \( \bar{I}^*(\omega) \) is a consistent estimator of the spectral density \( f_X(\omega) \) and \( \bar{I}^*(\omega) \) has the same asymptotic statistical properties of \( \bar{I}(\omega) \) given in Proposition 1.
The statistical properties of $\log \frac{f_X(\omega)}{f^*(\omega)}$, discussed previously, and Proposition 3 give the justification that both $\log \frac{f(\omega)}{f_X(\omega)}$ and $\log \frac{f^*(\omega)}{f_X(\omega)}$ have asymptotically the same distribution.

As previously mentioned, the asymptotic results presented here for the average bootstrap periodogram rely upon the fact that the process satisfies assumptions A1-A3, that is, it is a weakly stationary covariance process. As it is well-known, in this context the periodogram ordinates are asymptotically uncorrelated. However, the main interest in this paper is to use the proposed bootstrap methodology in the estimation of the fractional long-memory parameter in the ARFIMA($p, d, q$) processes (Section 4) which do not satisfy assumptions A1 and A3. Nevertheless, the empirical evidences presented in this paper support the use of the proposed methodology even in the case where, for fixed and small frequencies, the periodogram ordinates are not asymptotically uncorrelated.

4 The ARFIMA model and parameter estimation

Suppose now that the process $\{X_t\}$ has spectral density that behaves like

$$ f_X(\omega) \sim f^*(\omega) |\omega|^{-2d}, \text{ as } \omega \to 0 $$

where $f^*(\omega)$ is a spectral density that satisfies A3, and $d$ is a real number. When $d$ is positive, the assumption A1 does not hold and $\sum_{j=0}^{\infty} |\gamma(j)| = \infty$. A well-known parametric model with the above characteristics is the ARFIMA($p, d, q$) process which is the solution of the equation

$$ X_t - \mu = (1 - B)^{-d} \eta_t, \quad t \in \mathbb{Z}, $$

where $\eta_t = \phi(B)\theta(B)\epsilon_t$ is the ARMA($p, q$) process, $\mu$ is the mean (here it is assumed that $\mu = 0$), $\phi(B) = 1 - \sum_{j=1}^{p} \phi_j B^j$, $\theta(B) = 1 - \sum_{i=1}^{q} \theta_i B^i$, $p$ and $q$ are positive integers. $\Phi(z)$ and $\Theta(z)$, with a scalar $z$, are polynomials with all roots outside the unit circle and share no common factors. $d$ is the parameter that holds the memory of the process, that is, when $d \in (-0.5, 0.5)$ the ARFIMA($p, d, q$) process is said to be invertible and stationary. Besides, for $d \neq 0$ its autocovariance decays at a hyperbolic rate ($\gamma(j) = O(j^{-1+2d})$). For $d = 0$, $d \in (-0.5, 0)$ or $d \in (0, 0.5)$, the process is said to be short-memory, intermediate-memory or long-memory, respectively. The long-memory property is related to the behavior of the autocovariances which are not absolutely summable and the spectral density becomes unbounded at zero frequency. In the intermediate-memory region, the autocovariances are absolutely summable and, consequently, the spectral density is bounded.

The spectral density function of $\{X_t\}_{t \in \mathbb{Z}}$ is given by
Let $f_X(\omega) = f_\eta(\omega) \left[ 2 \sin \left( \frac{\omega}{2} \right) \right]^{-2d}$, $\omega \in [-\pi, \pi]$.

$f_X(\omega)$ is continuous except for $\omega = 0$ where it has a pole when $d > 0$. A recent review of the ARFIMA model and its properties can be found in Palma (2007) and Doukham et al. (2003).

Due to the singularity of $f_X(\omega)$ when $d > 0$, the standard results of the asymptotic distribution of the periodogram discussed previously can not be applied to $I_\omega(\omega_j)$ for small and fixed $j$. Hurvich and Beltrão (1993) showed that $\lim_{n \to \infty} \mathbb{E}\{ \frac{I_{X}(\omega_j)}{f_X(\omega_j)} \}$ depends on $j$ and $d$ and exceeds unity for most $d \neq 0$ (Künsch, 1986; Robinson, 1995). For $j \neq k$, $\frac{I_{X}(\omega_j)}{f_X(\omega_j)}$ and $\frac{I_{X}(\omega_k)}{f_X(\omega_k)}$ are correlated, and for a fixed value $j$ and Gaussian processes, the limiting distribution of $\frac{I_{X}(\omega_j)}{f_X(\omega_j)}$ is not exponential (Robinson, 1995). Not surprisingly, these properties are also observed for $\log \frac{I_{X}(\omega_j)}{f_X(\omega_j)}$. Hurvich and Beltrão (1993), Robinson (1995) and Hurvich et al. (1998) are some of the recent references that discuss deeply this matter and present, with theoretical rigor, results of the asymptotic properties of $\log \frac{I_{X}(\omega_j)}{f_X(\omega_j)}$ for fixed and small frequencies.

Concerning the averaged periodogram, for the ARFIMA process the following theorem gives the asymptotic distribution of $\frac{I_{X}(\omega_j)}{f_X(\omega_j)}$.

**Theorem 1.** Let $\bar{I}(\omega_j)$, $\omega_j \in (0, \pi)$, be the average of the periodogram of $r$ (fixed) independent partial realizations $x_{l,1}, ..., x_{l,n}$ of $\{X_l\}$, $l = 1, ..., r$, where $\{X_l\}$ is a Gaussian process generated by the stationary zero mean ARFIMA($p,d,q$) process with spectral function $f_X(\omega_j)$. The normalized average periodogram $\frac{I_{X}(\omega_j)}{f_X(\omega_j)}$ is asymptotically distributed as the quadratic form

$$\frac{\alpha_1}{2r} \chi_r^2 + \frac{\alpha_2}{2r} \chi_r^2$$

where $\alpha_1 = L_j(d) - 2L_j^*(d)$, $\alpha_2 = L_j(d) + 2L_j^*(d)$,

$$L_j(d) = \lim_{n \to \infty} \mathbb{E}\{ \frac{I_{X}(\omega_j)}{f_X(\omega_j)} \} = \frac{2}{\pi} \int_{-\infty}^{\infty} \frac{\sin^2(\omega/2)}{(2\pi j - \omega)^2} \left| \frac{\omega}{2\pi j} \right|^{-2d} d\omega$$

and

$$L_j^*(d) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\sin^2(\omega/2)}{(2\pi j - \omega)(2\pi j + \omega)} \left| \frac{\omega}{2\pi j} \right|^{-2d} d\omega.$$

Proof: Since $\frac{I_{X}(\omega_j)}{f_X(\omega_j)}$, $0 < j < n/2$, is a linear combination of $\frac{I_{X}(\omega_j)}{f_X(\omega_j)}$, the proof of Theorem 1 is straightforward obtained from Theorems 1 to 6 in Hurvich and Beltrão (1993).

Note that Proposition 1 is a particular case of the above theorem when $d = 0$. 


The asymptotic properties of \( \bar{I}(\omega_j) \), where the averaged periodogram is based on the division of the series into blocks, are given in Reisen et al. (2008). The authors show that the correlation of \( \bar{I}(\omega_j) \) evaluated at Fourier frequencies \( \omega_j \) and \( \omega_k \), for fixed \( j \) and \( k \), computed over different epochs does not vanish and the strength of the dependence among the epochs depends on \( d \). Nevertheless, under some additional conditions for \( f_\eta(\omega) \), they show that the bias is small for frequencies far away from zero and the dependence of the periodogram is small when the Fourier frequencies are sufficiently far apart. Their results are also translated to \( \log \bar{I}_x(\omega_j) \).

The main interest is to estimate the parameters of the process \( \{X_t\} \) that follows a Gaussian zero mean stationary ARFIMA(\( p, d, q \)) model. To estimate \( d \), this paper deals with a semiparametric estimation procedure based on the Ordinary Least Square (OLS) estimator proposed by Geweke and Porter-Hudak (1983) (GPH).

Since the GPH estimator is well-discussed in the literature, this method and its asymptotical statistical properties are briefly summarized here.

For a single realization \( x_1, ..., x_n \) of \( \{X_t\} \), the GPH estimate of \( d \) is obtained from the regression equation

\[
\log I_x(\omega) = a_0 - 2d \log [2 \sin(\omega/2)] + \xi, \tag{8}
\]

where the constant \( a_0 = \log f_\eta(0) + \log \frac{f_\eta(\omega)}{f_\eta(0)} + C, \xi = \log \frac{I_x(\omega)}{I_x(\omega)} - C \) and \( C = \phi(1) \).

The GPH estimate of \( d \) is given by

\[
d_{GPH} = (-0.5) \frac{\sum_{j=1}^{m} (v_j - \bar{v}) \log I_x(\omega_j)}{\sum_{j=1}^{m} (v_j - \bar{v})^2} \tag{9}
\]

where \( v_j = \log \{4 \sin^2(\omega_j/2)\} \) and \( m \) is the bandwidth in the regression equation which has to satisfies \( m \to \infty, n \to \infty \), with \( \frac{m}{n} \to 0 \) and \( \frac{m \log(m)}{n} \to 0 \).

Under these conditions, Hurvich et al. (1998) proved that the GPH-estimator is consistent for the memory parameter and asymptotically normal for Gaussian time series processes. The authors established that the optimal \( m \) in equations (8) and (9) is of order \( o(n^{4/5}) \) and \( (m)^{1/2}(d_{GPH} - d) \overset{d}{\to} N(0, \frac{\pi^2}{24}) \).

Reisen et al. (2008) propose using the average periodogram in equation (8), instead of a single periodogram, \( I_x(\omega) \), to estimate \( d \). The average periodogram can be obtained breaking the series in a sufficiently large number of epochs, calculating the periodogram in each epoch and averaging the periodograms. They prove, theoretically and empirically, that the mean square error of the estimates decreases as the number of epochs increases.

The difficulties concerning this approach is that the series can, sometimes, not be large enough to be broken in as many epochs as should be sufficient to reduce the mean square error. Thus, in this work the local bootstrap, discussed in the previous section, is used to obtain a large number of replications of the periodograms to calculate the averaging periodogram. This method then gives estimates of \( d \) by
replacing, in the above regression (Eq. 8), the periodogram $I_x(\omega_j)$ by the average local bootstrap $\bar{I}^*(\omega_j)$. This new approach will be called here GPH-LOC$_{\text{Aver}}$. As previously mentioned, the local bootstrap has also been used in time series with long-memory such as in Franco and Reisen (2004, 2007) and Arteche and Orbe (2008).

The method investigated by these authors is also considered here for comparison purpose. Basically, the estimation procedure consists in replacing the periodogram $I_x(\omega_j)$ by the local bootstrap periodogram $I^*(\omega_j)$, and the resulting estimator is called GPH-LOC.

5 Empirical results

The estimation methods discussed previously, the GPH estimator, the local bootstrap on the periodogram (GPH-LOC) and the local bootstrap on the average periodogram (GPH-LOC$_{\text{Aver}}$), are now investigated for finite sample sizes.

Series of sizes $n = 100, 1000$ from the ARFIMA$(1, d, 0)$ model (Eq. 6), with $d = 0.3$ and $\phi = 0, \pm 0.3$, were generated using the data generation method given in Hosking (1981), and the mean values, the sample variance (Var) and the mean square error (mse) were obtained over 1000 Monte Carlo replications.

Different values of the bandwidths $m$ in the regression equation and $\kappa_n$ in the probability $p_{\kappa_n,s} = \frac{1}{2\kappa_n+1}$ were considered with the aim to verify their finite sample properties in the estimates.

The finite empirical distribution of the standardized estimates were also obtained and compared with the standard normal distribution for a fixed confidence interval of level of 95%. For the percentile bootstrap confidence interval, 1000 bootstrap replications were taken. The empirical results are presented in the following tables.

The bootstrap approximations varied according to the number of periodogram ordinates $\kappa_n$, considered in the local bootstrap approach. According to Paparoditis and Politis (1999) increasing $\kappa_n$ increases the bias, while decreasing $\kappa_n$ increases the variance. Thus the values 1, 2 and 3 were chosen for $\kappa_n$ to study the effect of $\kappa_n$ on the performance of the estimators.

For the regression equation, the optimal bandwidth $m$ (see Hurvich et al. (1998)) was used in the model where $p = 1$, whereas $m = n/2$ for the ARFIMA$(0,d,0)$ model. The mean value, mean square error (mse) and variance of the Monte Carlo replications and the bootstrap approximations for the GPH are reported.

From the tables, it can be seen that the local average method (GPH-LOC$_{\text{Aver}}$) presents in general the smallest bias and mse, whatever the model chosen, a fact that is more evident for series of small size. Concerning the resampling width, $\kappa_n$, for the local bootstrap replicates, in general the smallest mse’s are obtained for $\kappa_n=3$. When an autoregressive term is present in the model, the parameter $d$ is underestimated for negative values of $\phi$ and overestimated for positive $\phi$’s. In these cases, especially for $\phi = 0.3$, the GPH-LOC$_{\text{Aver}}$ estimator presents a very good
performance for $\kappa_n=3$, with a very small bias and $mse$.

Table 1: Estimation results and empirical coverage rates for the ARFIMA(1,$d$,0) model, $n=100$

<table>
<thead>
<tr>
<th>$\kappa_n$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GPH</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.3053</td>
<td>0.3094</td>
<td>0.3082</td>
<td>0.1878</td>
<td>0.1979</td>
<td>0.1912</td>
<td>0.3964</td>
<td>0.4006</td>
<td>0.4083</td>
</tr>
<tr>
<td>MSE</td>
<td>0.0132</td>
<td>0.0139</td>
<td>0.0142</td>
<td>0.033</td>
<td>0.0306</td>
<td>0.0292</td>
<td>0.053</td>
<td>0.0534</td>
<td>0.0492</td>
</tr>
<tr>
<td>Var</td>
<td>0.0132</td>
<td>0.0138</td>
<td>0.0141</td>
<td>0.0204</td>
<td>0.0202</td>
<td>0.0174</td>
<td>0.0438</td>
<td>0.0433</td>
<td>0.0375</td>
</tr>
<tr>
<td><strong>GPH-LOC</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.3061</td>
<td>0.3116</td>
<td>0.3123</td>
<td>0.1899</td>
<td>0.1914</td>
<td>0.1846</td>
<td>0.4002</td>
<td>0.4108</td>
<td>0.4176</td>
</tr>
<tr>
<td>MSE</td>
<td>0.0164</td>
<td>0.0166</td>
<td>0.0159</td>
<td>0.0388</td>
<td>0.0343</td>
<td>0.0332</td>
<td>0.0668</td>
<td>0.062</td>
<td>0.0524</td>
</tr>
<tr>
<td>Var</td>
<td>0.0163</td>
<td>0.0164</td>
<td>0.0158</td>
<td>0.0253</td>
<td>0.0226</td>
<td>0.0198</td>
<td>0.0568</td>
<td>0.0497</td>
<td>0.0385</td>
</tr>
</tbody>
</table>

- **Mean** percentage in the confidence interval using the regression s.d.
- **GPH-LOC_Aver** percentage in the confidence interval using the asymptotic s.d.

<table>
<thead>
<tr>
<th>$\kappa_n$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GPH</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.2936</td>
<td>0.2984</td>
<td>0.3013</td>
<td>0.2686</td>
<td>0.2703</td>
<td>0.2722</td>
<td>0.317</td>
<td>0.3182</td>
<td>0.3213</td>
</tr>
<tr>
<td>MSE</td>
<td>0.0011</td>
<td>0.0012</td>
<td>0.0011</td>
<td>0.0034</td>
<td>0.0034</td>
<td>0.0033</td>
<td>0.0076</td>
<td>0.0074</td>
<td>0.0074</td>
</tr>
<tr>
<td>Var</td>
<td>0.0011</td>
<td>0.0012</td>
<td>0.0011</td>
<td>0.0024</td>
<td>0.0025</td>
<td>0.0025</td>
<td>0.0073</td>
<td>0.0071</td>
<td>0.0069</td>
</tr>
<tr>
<td><strong>GPH-LOC</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.2904</td>
<td>0.2922</td>
<td>0.301</td>
<td>0.2671</td>
<td>0.2704</td>
<td>0.2728</td>
<td>0.3467</td>
<td>0.3515</td>
<td>0.3526</td>
</tr>
<tr>
<td>MSE</td>
<td>0.0013</td>
<td>0.0014</td>
<td>0.0012</td>
<td>0.0041</td>
<td>0.0039</td>
<td>0.0035</td>
<td>0.0086</td>
<td>0.0086</td>
<td>0.0082</td>
</tr>
<tr>
<td>Var</td>
<td>0.0013</td>
<td>0.0014</td>
<td>0.0012</td>
<td>0.0031</td>
<td>0.0031</td>
<td>0.0027</td>
<td>0.0083</td>
<td>0.0082</td>
<td>0.0078</td>
</tr>
<tr>
<td><strong>GPH-LOC_Aver</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.2944</td>
<td>0.2914</td>
<td>0.3004</td>
<td>0.2666</td>
<td>0.2699</td>
<td>0.2719</td>
<td>0.3442</td>
<td>0.3414</td>
<td>0.3429</td>
</tr>
<tr>
<td>MSE</td>
<td>0.0001</td>
<td>0.0009</td>
<td>0.0008</td>
<td>0.0035</td>
<td>0.0035</td>
<td>0.0035</td>
<td>0.0063</td>
<td>0.0058</td>
<td>0.0054</td>
</tr>
<tr>
<td>Var</td>
<td>0.0001</td>
<td>0.0009</td>
<td>0.0008</td>
<td>0.0023</td>
<td>0.0023</td>
<td>0.0023</td>
<td>0.0057</td>
<td>0.0057</td>
<td>0.0052</td>
</tr>
</tbody>
</table>

- **GPH-LOC_Aver** percentage in the confidence interval using the regression s.d.
- **GPH-LOC_Aver** percentage in the confidence interval using the asymptotic s.d.

For the ARFIMA(0,$d$,0) model, the asymptotic confidence interval using the asymptotic variance presents coverage slightly closer to the nominal level of 0.95 than the asymptotic interval using the variance estimated from the regression equation. Tables also shows that the asymptotic intervals are affected by autoregressive terms, as the coverage is farther form 0.95 in these cases.

When $m = n/2$, it can be also verified that the GPH-LOC_Aver method presents the smallest $mse$, whatever the model chosen. However, in general the best results are obtained for $\kappa_n=2$. Once more, for the ARFIMA(0,$d$,0) model, the parameter $d$ is underestimated for negative values of $\phi$ and overestimated for positive $\phi$'s, but this
time with a smaller bias and larger $mse$ compared to the optimal bandwidth. The performance of the asymptotic confidence intervals are the same as the case of the optimal bandwidth. As expected, increasing the sample size reduces substantially the bias and $mse$.

Figures 1a to 1d illustrate graphically some empirical distributions of the standardized estimates of the models in the above tables and the $N(0,1)$ density curve. These pictures clearly evidence the percentage results given in the tables.

![Empirical distributions](image)

(a) $n = 100$, $p = 0$, $k_n = 1$

(b) $n = 100$, $p = 1$, $\phi = 0.3$, $k_n = 1$

(c) $n = 1000$, $p = 0$, $k_n = 1$

(d) $n = 1000$, $p = 1$, $\phi = 0.3$, $k_n = 1$

Figure 1: Empirical distributions of the standardized estimates of the models and the $N(0,1)$ density curve.
6 Concluding remarks

In this paper, a new estimator for the long memory parameter of ARFIMA models has been proposed, using the bootstrap. The estimator is a semiparametric approach based on the log average periodogram, following the idea of Reisen et al. (2008) of using least square methods to calculate $d$. To obtain replications of the periodogram in order to calculate the average periodogram, the local bootstrap (Paparoditis and Politis, 1999, and Silva et al., 2006) has been used. The average periodogram thus obtained is shown to provide an unbiased and consistent estimator of the spectral density.

A Monte Carlo study to verify the finite sample properties of the proposed estimator has also been performed. The simulations have shown that the new estimator presents smaller mean square errors than the Geweke and Porter-Hudak (1983) estimator ($GPH$) and than the local bootstrap applied to the $GPH$.

Acknowledgements

The author V.A. Reisen gratefully acknowledges partial financial support from CNPq/ Brazil. G.C. Franco was partially supported by CNPq-Brazil, and also by Fundação de Amparo à Pesquisa no Estado de Minas Gerais (FAPEMIG Foundation). Giovanni Comarela thanks the grant from PIBIC-CNPq. All the authors thank the undergraduate student Pedro Berger (PIBIC-CNPq) for his help in some simulation works.

References


