Computational Statistics and Data Analysis 58 (2013) 242–256

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

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

Minimum distance estimation of ARFIMA processes

Mauricio Zevallos [a,*], Wilfredo Palma [b]

[a] Department of Statistics, University of Campinas, IMECC-UNICAMP, CEP 13083-859, Campinas, SP, Brazil

[b] Department of Statistics, Pontificia Universidad Católica de Chile, Chile

ARTICLE INFO

Article history:
Received 29 September 2010
Received in revised form 6 August 2012
Accepted 6 August 2012
Available online 20 August 2012

Keywords:
Autocorrelation
Fractional noise
Fractional filtering
Long-memory
Missing data
Non-Gaussian processes

ABSTRACT

This paper proposes a new minimum distance methodology for the estimation of ARFIMA processes with Gaussian and non-Gaussian errors. The main advantage of this method is that it allows for a computationally efficient estimation when the long-memory parameter is in the interval $d \in (-\frac{1}{2}, \frac{1}{2})$. Previous minimum distance estimation techniques are usually limited to the range $d \in (-\frac{1}{2}, \frac{1}{4})$, leaving outside the very important case of strong long memory with $d \in [\frac{1}{4}, \frac{1}{2})$. It is shown that the new estimator satisfies a central limit theorem and Monte Carlo experiments indicate that the proposed estimator performs very well even for small sample sizes. The methodology is illustrated with three applications. The first two examples involve real-life time series while the third application illustrates that the proposed methodology is a sound alternative for dealing with incomplete time series.

1. Introduction

Long-memory time series are characterized by significant autocorrelations at large lags. One of the most well known models for dealing with this type of data is the autoregressive fractionally integrated moving average (ARFIMA) processes proposed by Granger and Joyeux (1980) and Hosking (1981). These processes have been widely used for modeling long-range dependence; see for example Doukhan et al. (2003), Bhardwaj and Swanson (2006) and Palma (2007), among others.

In recent years, the problem of estimating long-memory models has been extensively discussed. For Gaussian processes, two exact maximum likelihood estimation algorithms have been proposed. The first approach, proposed by Sowell (1992), is computationally demanding and could be numerically unstable for certain values of the Hurst parameter. However, this method can be improved by using appropriate numerical devices, such as those discussed by Doornik and Ooms (2003). The second method is based on a state-space representation of the process and the estimation is carried out by Kalman filters. This approach, discussed by Chan and Palma (1998) allows for the treatment of missing data. Furthermore, several other approximate likelihood methods have been proposed in the literature, including Haslett and Raftery (1989) and Cheang and Reinsel (2003) in the time domain, and many semiparametric estimators in the frequency domain. For an account of these methodologies, see Robinson (1995) and Henry (2001) for estimation of nonlinear long-memory processes, and Liseo et al. (2001) for Bayesian estimation techniques.

Tieslau et al. (1996) proposed the so-called *minimum distance estimator* (MDE) as an alternative for estimating *fractional noise* processes; see their definition in the next section. The MDE method is based on the minimization of the distance between sample and population autocorrelations. An advantage of this technique over other approaches is its high numerical efficiency. However, a disadvantage of this MDE technique is that the estimate of the long-memory parameter d turns out to

* Corresponding author.
E-mail address: amadeus@ime.unicamp.br (M. Zevallos).

0167-9473/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2012.08.005

be consistent and asymptotically normal only for $d \in (-\frac{1}{2}, \frac{1}{4})$. Therefore, this approach has limited application in practice because it is not known *a priori* whether d belongs to $(-\frac{1}{2}, \frac{1}{4})$ or $[\frac{1}{4}, \frac{1}{2})$.

In order to obtain an asymptotically normal estimator of the long-memory parameter d for the entire region of stationarity, Mayoral (2007) proposed a MDE procedure based on the squared residuals obtained after filtering a series through an ARFIMA model. This idea was later employed by Kouamé and Hili (2008) for estimating Gegenbauer autoregressive moving average (GARMA) processes. Additionally, autocorrelations of squared residuals were considered by Baillie and Chung (2001) for obtaining minimum distance estimators of GARCH models.

Instead of using squared observations or squared residuals, in this paper we propose a new MDE method based on the fractional filtering of the series. The main features of these estimators, called MDEFF hereafter, can be summarized as follows. They are easy to calculate and numerically efficient since the computation of the MDEFF estimates is based on a reduced number of sample autocorrelations. This MDEFF approach allows for the estimation of Gaussian and non-Gaussian ARFIMA processes since it does not depend on the distribution of the time series data. The estimator of d is not affected by the estimation of the level parameter μ since it is based only on the sample autocorrelation function. As shown by Cheung and Diebold (1994), this does not occur for the exact maximum likelihood estimator (MLE) implemented by Sowell (1992). On the other hand, several Monte Carlo experiments evidence that the MDEFF method displays small bias and its precision, measured in terms of the mean squared error, is very good. For some ARFIMA models, the MDEFF technique exhibits a substantially better performance than other widely used methods. The MDEFF approach works well even with missing data. Furthermore, given their numerical efficiency, the MDEFF approach is useful for handling huge datasets, typically found in the context of long-memory time series.

On the other hand, a key component of the MDE methodology is the asymptotic variance of sample autocorrelations. An additional contribution of this paper is an analytic expression for this asymptotic variance, provided in Section 2.

The remainder of this paper is organized as follows. In Section 2 we briefly review the large-sample behavior of the sample autocorrelations of a long-memory process. In particular, we establish a useful formula for the calculation of the asymptotic distribution of the sample autocorrelation function. A new MDE method based on minimum distance and fractional filtering is proposed in Section 3. The results from several Monte Carlo simulations assessing the finite sample performance of this method are reported in Section 4. Section 5 is devoted to the application of the proposed methodology to the estimation of two well known long-memory time series. In addition, this section presents a study of the effectiveness of MDEFF method for parameter estimation in the context of missing data. Final remarks are given in Section 6 while proofs of the results established in this paper are provided in the technical appendix.

2. Asymptotic distribution of sample autocorrelations

Let $\{y_1, \dots, y_n\}$ be a sequence of observations from the ARFIMA(p, d, q) process defined by

$$\phi(B)y_t = \theta(B)(1-B)^{-d}\varepsilon_t, \tag{1}$$

where B is the backshift operator $By_t = y_{t-1}$, $(1-B)^{-d}$ is the fractional difference operator defined by the binomial series $(1-B)^{-d} = \sum_{k=0}^{\infty} \frac{(k+d-1)!}{k!(d-1)!} B^k$, and $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$, $\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q$. It is assumed that $\{\varepsilon_t\}$ is a zero-mean independent identically distributed sequence with $E(\varepsilon_t^2) = \sigma^2$ and $E(\varepsilon_t^4) < \infty$.

Let ρ and $\widehat{\rho}$ be two vectors containing the first M population and sample autocorrelations, respectively

$$\rho = (\rho_1, \dots, \rho_M)', \qquad \widehat{\rho} = (\widehat{\rho}_1, \dots, \widehat{\rho}_M)'.$$

Hosking (1996) showed that for an ARFIMA(p, d, q) process with $d \in (-\frac{1}{2}, \frac{1}{4})$ and parameter vector $\lambda = (\phi_1, \dots, \phi_p, d, \theta_1, \dots, \theta_q)$, the vector of estimated autocorrelations satisfies the central limit theorem (CLT)

$$\sqrt{n}\,(\widehat{\rho} - \rho) \to N[0, V(\lambda)], \tag{2}$$

for fixed M, as $n \to \infty$, where the elements $V(\lambda)$ are given by

$$V(\lambda)_{ij} = \sum_{k=-\infty}^{\infty} \left[\rho(k+i)\rho(k+j) + \rho(k-i)\rho(k+j) + 2\rho(i)\rho(j)\rho^2(k) - 2\rho(i)\rho(k)\rho(k+j) - 2\rho(j)\rho(k)\rho(k+i)\right]. \tag{3}$$

Observe that the result (2) can be extended to $d \in (-\infty, \frac{1}{4})$ by an application of the CLT of Hannan (1976). In this interval for d, the spectral density of $\{y_t\}$ is squared integrable, which is a necessary and sufficient condition for (2) according to Hannan (1976); see the proof of Theorem 1 for further details. On the other hand, note that expression (3) can be written as

$$V(\lambda)_{ij} = \varphi(j-i) + \varphi(j+i) + 2\rho(i)\rho(j)\varphi(0) - 2\rho(i)\varphi(j) - 2\rho(j)\varphi(i), \tag{4}$$

where

$$\varphi(k) = \sum_{j=-\infty}^{\infty} \rho(j)\rho(j+k). \tag{5}$$

The next proposition, proved in the technical appendix, provides a useful expression to calculate the function $\varphi(\cdot)$.

Proposition 1. *Let $\{y_t\}$ be an* ARFIMA(p, d, q) *process defined in (1) with $d < \frac{1}{4}$ and autocovariance function $\gamma(\cdot)$. Assume that the roots $\{r_1, \ldots, r_p\}$ of the AR polynomial are simple. Then,*

$$\sum_{j=-\infty}^{\infty} \gamma(j)\gamma(j+s) = \sigma^4 \sum_{k=-2q}^{2q} \sum_{j=1}^{p} \psi(k) C(d, 2p+k+s, r_j), \tag{6}$$

where

$$\begin{aligned} C(d, h, r_j) \;=\; & \frac{\Gamma(1-4d)}{\Gamma(1-2d)\Gamma(2d)} \frac{\Gamma(h+2d)}{\Gamma(1+h-2d)} c_j \left\{ \alpha_j F(h+2d, 1; 1+h-2d; r_j) \right. \\ & + \beta_j [1 - F(2d-h, 1; 1-h-2d; r_j)] + r_j^{4p-2} F(h+2d, 2; h+1-2d; r_j) \\ & \left. + r_j^{-2} [F(2d-h, 2; 1-h-2d; r_j) - 2F(2d-h, 1; 1-h-2d; r_j) + 1] \right\}, \end{aligned}$$

$$\alpha_j = -2 r_j^{4p-2} \left[\sum_{k=1}^{p} \frac{1}{1 - r_j r_k} + \sum_{k \neq j} \frac{r_k}{r_k - r_j} \right],$$

$$\beta_j = -\frac{2}{r_j} \left[\sum_{k=1}^{p} \frac{r_k}{1 - r_j r_k} + \sum_{k \neq j} \frac{1}{r_k - r_j} \right],$$

$$c_j = \prod_{k=1}^{p} (1 - r_j r_k)^{-2} \Pi_{k \neq j} (r_k - r_j)^{-2},$$

$$\psi(k) = \sum_{i,j,\ell=0}^{q} \theta_i \theta_j \theta_\ell \theta_{\ell+j-i+k},$$

with $\theta_0 = 1$, $\theta_j = 0$ for $j < 0$ and $F(\cdot)$ is the hypergeometric function.

From (6) we can readily obtain the expression

$$\varphi(k) = \frac{1}{[\gamma(0)]^2} \sum_{j=-\infty}^{\infty} \gamma(j)\gamma(j+k),$$

where $\gamma(0)$ is given by formula (8) of Sowell (1992). For a model with no autoregressive components, Proposition 1 provides a simple expression for $\varphi(\cdot)$. For example, in the case of an ARFIMA$(0, d, 1)$ model we have the following result which is an immediate consequence of Proposition 1.

Corollary 1. *Let $\{y_t\}$ be an* ARFIMA$(0, d, 1)$ *process satisfying*

$$y_t = (1 + \theta B)(1 - B)^{-d} \varepsilon_t,$$

with $|\theta| < 1$ and $d < \frac{1}{4}$. Then, (5) can be expressed as

$$\varphi(k) = \frac{1}{\beta^2} \frac{\Gamma(1-4d)\Gamma^4(1-d)}{\Gamma^2(1-2d)} \frac{(-1)^k}{\Gamma(1-2d+k)\Gamma(1-2d-k)} s_k,$$

where $\beta = \theta^2 + 2\theta d/(1-d) + 1$ and for $k = 0, 1, \ldots$

$$s_k = (1+\theta)^4 + a_k \theta \left[\theta b_k - 2(\theta+1)^2 \right],$$

$$a_k = \frac{(1-4d)(2-4d)}{(1-2d+k)(1-2d-k)},$$

$$b_k = \frac{(3-4d)(4-4d)}{(2-2d+k)(2-2d-k)}.$$

Note that for an ARFIMA$(0, d, 0)$ model we have $s_k = 1$ for all k and $\beta = 1$. Consequently, for this model we have

$$\varphi(k) = \frac{\Gamma(1-4d)\Gamma^4(1-d)}{\Gamma^2(1-2d)} \frac{(-1)^k}{\Gamma(1-2d+k)\Gamma(1-2d-k)}, \tag{7}$$

which can be easily computed; see Section 3.2.1. In addition, by setting $d = 0$, Proposition 1 can also be applied to the ARMA$(1, 1)$ case discussed by Baillie and Chung (2001).

3. Minimum distance estimation

The MDE method is based on the minimization of the distance between sample and population autocorrelations. This approach takes advantage of the good statistical properties displayed by sample correlations when a central limit theorem (2) holds. The MDE of λ is the value that minimizes the criterion function

$$S(\lambda) = [\rho(\lambda) - \widehat{\rho}\,]'V(\lambda)^{-1}[\rho(\lambda) - \widehat{\rho}\,], \tag{8}$$

where $\rho(\lambda)$ is the true autocorrelation function (ACF) for the parameter λ. Let $\widehat{\lambda}$ be the value that minimizes $S(\lambda)$ and let λ_0 be the true parameter. The MDE method intends to find conditions so that

$$\sqrt{n}(\widehat{\lambda} - \lambda_0) \to N[0, \Lambda(\lambda)], \tag{9}$$

as $n \to \infty$, where the variance–covariance matrix $\Lambda(\cdot)$ is given by

$$\Lambda(\lambda) = \left[D(\lambda)'V(\lambda)^{-1}D(\lambda)\right]^{-1}, \tag{10}$$

with $D(\lambda) = \partial\rho(\lambda)/\partial\lambda$. The limiting distribution (9) holds for the MDE method proposed by Tieslau et al. (1996) for a fractional noise process with long-memory parameter satisfying $d \in (-\frac{1}{2}, \frac{1}{4})$. In what follows, we briefly review this result and propose a new MDE method applicable to the entire class of ARFIMA(p, d, q) processes with $d \in (-\frac{1}{2}, \frac{1}{2})$.

3.1. *Fractional noise estimation when* $-\frac{1}{2} < d < \frac{1}{4}$

As shown by Hosking (1996), when $d \in (-\frac{1}{2}, \frac{1}{4})$, the sample autocorrelations of an ARFIMA model satisfy a central limit theorem with a rate $n^{-1/2}$ as in (2). For ARFIMA$(0, d, 0)$ models with $d \in (-\frac{1}{2}, \frac{1}{4})$, Tieslau et al. (1996) proved (9)–(10) considering $\lambda = d$, where $D(d) = (D_1, \dots, D_M)'$ is given by

$$D_i = \rho_i \sum_{j=1}^{i} \frac{2j-1}{(j-1+d)(j-d)}.$$

Unfortunately, the method proposed by Tieslau et al. (1996) requires that $d < \frac{1}{4}$, but the worst effects on the statistical inferences about a long-range dependent time series occur for large values of d or *very long memory*, that is, when $\frac{1}{4} < d < \frac{1}{2}$; see Section 1.1 of Beran (1994). For this interval, as pointed out by Hosking (1996), the asymptotic distribution of sample autocorrelations is non-Gaussian and the convergence rate depends on the Hurst parameter, which is unknown in practical applications.

3.2. *ARFIMA estimation with* $-\frac{1}{2} < d < \frac{1}{2}$

In order to overcome the problem described in the previous section, our strategy to handle strongly dependent processes with *very long memory* consists of applying a fractional filter to the observed series to reduce its memory. Assume that $\{y_t\}$ is an ARFIMA(p, d, q) process defined by (1). Then, by applying the fractional filter $(1 - B)^{d_0}$ to $\{y_t\}$ we obtain the process $\{x_t\}$ given by

$$x_t = (1 - B)^{d_0} y_t = \sum_{j=0}^{\infty} \pi_j(d_0) y_{t-j}, \tag{11}$$

where the coefficients $\pi_j(d_0)$ can be readily evaluated by the recursive equation

$$\pi_{j+1}(d_0) = \pi_j(d_0)\,\frac{j - d_0}{j + 1},$$

for $j = 0, 1, \dots,$ with $\pi_0(d_0) = 1$. Thus, the filtered process $\{x_t\}$ satisfies the discrete-time equation

$$\phi(B)x_t = \theta(B)(1 - B)^{d_0 - d}\varepsilon_t.$$

Consequently, $\{x_t\}$ is an ARFIMA$(p, d - d_0, q)$ process and its memory is reduced by choosing $d_0 > 0$. As an example, assume that $d_0 = 0.5$. Thus, if $d \in (-\frac{1}{2}, \frac{1}{2})$ then the memory parameter of the filtered process $\{x_t\}$ lies in the interval $(-1, 0)$. Hence, according to Bondon and Palma (2007), the process $\{x_t\}$ is stationary and invertible. The reduction of memory implies that the sample autocorrelations of $\{x_t\}$ satisfy the central limit theorem (2). Then, the MDE of an ARFIMA model can be calculated on the filtered series. This procedure corresponds to the MDEFF method introduced in Section 1. In practice, the observed time series $\{y_1, \dots, y_n\}$ is a finite set of values but in order to apply the fractional filter (11) we need the entire infinite series $\{y_n, y_{n-1}, \dots\}$. To circumvent this problem, the filtered series can be approximated by

$$\tilde{x}_t = \sum_{j=0}^{t-1} \pi_j(d_0) y_{t-j}, \quad t = 1, \dots, n. \tag{12}$$

The procedure for estimating the parameters is summarized next. For illustration purposes, we consider an ARFIMA(0, d, 1) process with parameter $\lambda = (d, \theta)$: (a) Apply the fractional filter (12) using $d_0 = 0.5$ to the demeaned observed series $\{y_1, \ldots, y_n\}$ to obtain the filtered series $\{\tilde{x}_1, \ldots, \tilde{x}_n\}$. (b) Calculate the autocorrelations $\widehat{\rho}_{\tilde{x}}(1), \ldots, \widehat{\rho}_{\tilde{x}}(M)$ based on $\{\tilde{x}_1, \ldots, \tilde{x}_n\}$ and find the minimum distance estimate by minimizing the criterion function $S(\cdot)$ defined in (8). These estimates are denoted by $(\widehat{d}_f, \widehat{\theta})$. Therefore the MDEFF of $\{y_1, \ldots, y_n\}$ is $\widehat{\lambda} = (d_0 + \widehat{d}_f, \widehat{\theta})$. (c) The estimated standard deviations (SD) are computed from the matrix $\Lambda(\widehat{\lambda})$ defined in (10). For instance, the SD of $\widehat{d}$ and $\widehat{\theta}$ are given by

$$\mathrm{SD}(\widehat{d}) = \left[n^{-1}\Lambda_d(\widehat{d}_f, \widehat{\theta})\right]^{\frac{1}{2}}, \qquad \mathrm{SD}(\widehat{\theta}) = \left[n^{-1}\Lambda_\theta(\widehat{d}_f, \widehat{\theta})\right]^{\frac{1}{2}}, \tag{13}$$

where Λ_d and Λ_θ are the diagonal elements of Λ corresponding to d and θ, respectively. A formal justification of this procedure is given in the next theorem, which is proved in the technical appendix.

Theorem 1. *Let $\{y_t\}$ be an ARFIMA(p, d, q) process with $-\frac{1}{2} < d < \frac{1}{2}$ and $-1 < d - d_0 < \frac{1}{4}$, for some $d_0 > 0$. Let $\lambda = (\phi_1, \ldots, \phi_p, d, \theta_1, \ldots, \theta_q)$ be the true parameter vector of this model. In addition, let $\{x_t\}$ be the process given in (11) obtained by fractional filtering and let $\{\tilde{x}_t\}$ be its finite sample counterpart given by (12). Let $\widehat{\rho}_x = [\widehat{\rho}_x(1), \ldots, \widehat{\rho}_x(M)]'$ and $\widehat{\rho}_{\tilde{x}} = [\widehat{\rho}_{\tilde{x}}(1), \ldots, \widehat{\rho}_{\tilde{x}}(M)]'$ be the sample autocorrelation vectors of $\{x_t\}$ and $\{\tilde{x}_t\}$, respectively. Then, for fixed M, the autocorrelation vector $\widehat{\rho}_{\tilde{x}}$ satisfies*

$$\sqrt{n}(\widehat{\rho}_{\tilde{x}} - \rho_x) \to N[0, V(\lambda)],$$

as $n \to \infty$, where $V(\lambda)$ is the matrix defined in (3). Furthermore, the MDEFF estimator $\widehat{\lambda}$ satisfies

$$\sqrt{n}(\widehat{\lambda} - \lambda) \to N[0, \Lambda(\lambda^*)],$$

as $n \to \infty$, where $\Lambda(\cdot)$ is the matrix defined in (10) evaluated at $\lambda^* = (\phi_1, \ldots, \phi_p, d - d_0, \theta_1, \ldots, \theta_q)$.

Remark 1. Note that by choosing $d_0 = \frac{1}{2}$ the assumption $-1 < d - d_0 < \frac{1}{4}$ in Theorem 1 is satisfied for all $d \in (-\frac{1}{2}, \frac{1}{2})$.

It is worthwhile comparing the theoretical efficiency of the MDEFF and the exact Gaussian maximum likelihood estimation methods. For instance, Table 1 displays the ratios of the standard deviations, MDEFF over MLE, for several ARFIMA(0, d, 1) models. The standard deviations of the MDEFF method are calculated considering $M = 10, 20, 50$ and by using formulas (13) with theoretical parameter values instead of estimates. For a Gaussian fractional noise time series of size n, the asymptotic standard deviation of the maximum likelihood estimate of d is $\sqrt{6/(n\pi^2)}$. According to Palma (2007, p. 106), for a Gaussian ARFIMA(0, d, 1) time series we have

$$\mathrm{SD}(\widehat{d}) = \frac{1}{\pi}\left[\frac{6}{n(1-\rho_\theta^2)}\right]^{\frac{1}{2}}, \qquad \mathrm{SD}(\widehat{\theta}) = \frac{1}{\pi}\left[\frac{1-\theta^2}{n(1-\rho_\theta^2)}\right]^{\frac{1}{2}},$$

where $\rho_\theta = -\frac{\sqrt{6}}{\pi}\sqrt{1-\theta^2}\,\frac{\log(1+\theta)}{\theta}$ is the correlation between $\widehat{d}$ and $\widehat{\theta}$.

Table 1 reveals that the efficiency of the MDEFF method increases with M. In particular, for $M = 50$ the efficiency of the MDEFF estimates seems to be very good. For $M = 10$ and $M = 20$ the efficiency seems to be good when $\theta \geq 0$ and $d \geq 0.30$ and reasonable when $d < 0.30$. Overall, these results indicate that for $\theta \geq 0$ the efficiency of the MDEFF method is good even for $M = 10$ autocorrelations. In Section 4, the performance of MDEFF is evaluated in finite samples through Monte Carlo simulations which show that the performance of the MDEFF estimator is very good compared to other techniques.

3.2.1. Implementation of the MDEFF method

The implementation of the MDEFF method is illustrated in this section by considering the ARFIMA(0, d, 1) model with $\lambda = (d, \theta)$. In this example, we obtain the following expressions, which include the fractional noise process as a particular case when $\beta = 1$ and $s_k = 1$ for all k. These quantities have been defined in Corollary 1. (a) *Autocorrelation function*: By Hosking (1981) we have

$$\rho(k) = \frac{\Gamma(k+d)\Gamma(1-d)}{\Gamma(k-d+1)\Gamma(d)}\left[\frac{ak^2-(1-d)^2}{k^2-(1-d)^2}\right],$$

where $a = (1+\theta)^2/[1+\theta^2+2\theta d/(1-d)]$. For a fractional noise model, $\rho(1) = d/(1-d)$ and the autocorrelations can be calculated recursively by the formula,

$$\rho(k+1) = \rho(k)\frac{k+d}{k-d+1}, \quad k = 0, 1, \ldots, M-1.$$

(b) *Calculation of* $V(\lambda)$: Consider $\varphi(k) = Cf(k)$ in (4) where

$$C = \frac{1}{\beta^2}\frac{\Gamma(1-4d)\Gamma^4(1-d)}{\Gamma^4(1-2d)}, \qquad f(k) = \frac{(-1)^k\Gamma^2(1-2d)}{\Gamma(1-2d+k)\Gamma(1-2d-k)}s_k.$$

Table 1
Efficiency of MDEFF based on M autocorrelations and $d_0 = 0.5$ for ARFIMA(0, d, 1) and fractional noise (FN) models. R_d and R_θ correspond to the ratio of the asymptotic MDEFF standard deviation over the asymptotic MLE standard deviation for d and θ, respectively.

d	θ	$M = 10$		$M = 20$		$M = 50$	
		R_d	R_θ	R_d	R_θ	R_d	R_θ
0.45	0.8	1.061	1.154	1.030	1.008	1.014	1.001
	0.4	1.075	1.024	1.041	1.013	1.019	1.006
	FN	1.052		1.028		1.013	
	−0.4	1.398	1.322	1.166	1.133	1.066	1.052
	−0.8	1.216	1.245	1.170	1.131	1.083	1.068
0.40	0.8	1.085	1.154	1.046	1.009	1.022	1.002
	0.4	1.105	1.031	1.061	1.018	1.029	1.009
	FN	1.076		1.043		1.020	
	−0.4	1.519	1.405	1.226	1.175	1.094	1.073
	−0.8	1.223	1.269	1.161	1.115	1.094	1.072
0.30	0.8	1.136	1.152	1.078	1.011	1.040	1.003
	0.4	1.169	1.046	1.101	1.028	1.051	1.014
	FN	1.125		1.074		1.037	
	−0.4	1.785	1.587	1.353	1.263	1.155	1.116
	−0.8	1.245	1.341	1.147	1.097	1.109	1.076
0.20	0.8	1.186	1.151	1.111	1.013	1.058	1.005
	0.4	1.234	1.060	1.142	1.037	1.073	1.020
	FN	1.175		1.105		1.054	
	−0.4	2.085	1.791	1.487	1.355	1.217	1.159
	−0.8	1.276	1.435	1.140	1.095	1.118	1.076
0.10	0.8	1.236	1.150	1.143	1.015	1.075	1.006
	0.4	1.298	1.074	1.183	1.046	1.095	1.025
	FN	1.224		1.136		1.070	
	−0.4	2.422	2.017	1.628	1.450	1.279	1.202
	−0.8	1.313	1.546	1.139	1.105	1.122	1.074

Then, for $i = 1, \ldots, M,\ j = i, \ldots, M$ the elements of matrix $V(\lambda)$ are

$$V(\lambda)_{ij} = C\left[f(j-i) + f(j+i) + 2\rho(i)\rho(j)f(0) - 2\rho(i)f(j) - 2\rho(j)f(i)\right].$$

For a fractional noise model, $\beta = 1, f(0) = 1$. Hence, a simplified formula for calculating $f(\cdot)$ is given by,

$$f(k+1) = f(k)\frac{2d+k}{1-2d+k}, \quad k = 0, 1, \ldots, 2M-1.$$

(c) *Computation of* $D(\lambda)$: The first row of this matrix is $\frac{\partial\rho}{\partial d} = [D_d(1), \ldots, D_d(M)]$ and the second row is $\frac{\partial\rho}{\partial\theta} = [D_\theta(1), \ldots, D_\theta(M)]$, where

$$D_d(k) = \rho(k)\left[T_k + \sum_{j=1}^{k}\frac{2j-1}{(j-1+d)(j-d)}\right],$$

$$T_k = \frac{-2(1-d)}{k^2-(1-d)^2} + \frac{\left(\frac{\partial a}{\partial d}\right)k^2 + 2(1-d)}{ak^2-(1-d)^2},$$

$$\frac{\partial a}{\partial d} = \frac{-2\theta(1+\theta)^2}{\left[(1+\theta^2)(1-d)+2\theta d\right]^2},$$

and

$$D_\theta(k) = \rho(k)\left(\frac{\partial a}{\partial\theta}\right)\left[\frac{k^2}{ak^2-(1-d)^2}\right],$$

$$\frac{\partial a}{\partial\theta} = 2a\left[\frac{1}{1+\theta} - \frac{\theta(1-d)+d}{(1+\theta^2)(1-d)+2\theta d}\right].$$

3.2.2. MDEFF for time series with missing data

The MDEFF method can be easily adapted to handle time series with missing data because the autocorrelations can be estimated despite the presence of missing data; see for example Parzen (1963). A minor difficulty arises in the filtering step (12) since the calculation of $\tilde{x}_t$ requires all the observations $\{y_t, \ldots, y_1\}$. A solution for this problem is interpolating the missing values by using, for instance, cubic splines. Thus, let $y_1, \ldots, y_n$ be the available time series with missing data at positions indexed by a set A. The steps for the application of the MDEFF method to this case are: (a) Interpolate the

Table 2
Theoretical SD of the MDEFF of fractional noise series of size n, based on $M = 10$ autocorrelations and $d_0 = 0.5$. The asymptotic SD of the MLE is reported in the last row.

d	$n = 100$	$n = 250$	$n = 500$
0.45	0.0820	0.0519	0.0367
0.40	0.0839	0.0530	0.0375
0.20	0.0916	0.0579	0.0410
MLE	0.0780	0.0493	0.0349

Table 3
Theoretical SD for MDEFF of ARFIMA(0, d, 1) time series of size n with $d = 0.4$ based on $M = 10$ autocorrelations and $d_0 = 0.5$. Asymptotic standard deviations for the MLE are also included.

θ	Method	$n = 100$		$n = 250$		$n = 500$	
		SD(d)	SD(θ)	SD(d)	SD(θ)	SD(d)	SD(θ)
0.8	MDEFF	0.0901	0.0737	0.0570	0.0466	0.0403	0.0330
	MLE	0.0830	0.0639	0.0525	0.0404	0.0371	0.0286
0.4	MDEFF	0.1078	0.1183	0.0682	0.0748	0.0482	0.0529
	MLE	0.0976	0.1147	0.0617	0.0725	0.0436	0.0513
−0.4	MDEFF	0.2896	0.3150	0.1832	0.1992	0.1295	0.1409
	MLE	0.1907	0.2242	0.1206	0.1418	0.0853	0.1002
−0.8	MDEFF	0.2820	0.2253	0.1784	0.1425	0.1261	0.1007
	MLE	0.2307	0.1775	0.1459	0.1123	0.1032	0.0794

missing data A by using splines. Denote the resultant series as $\tilde{y}_1, \ldots, \tilde{y}_n$. (b) Obtain the filtered series through (12) by using $\tilde{y}_1, \ldots, \tilde{y}_n$ instead of $y_1, \ldots, y_n$. (c) Calculate the autocorrelations on $\tilde{x}_1, \ldots, \tilde{x}_n$ considering the missing data indexed by A. (d) Apply the MDEFF method as described in Section 3.2. Note that the interpolation by cubic splines is simply a tool to avoid missing more observations when filtering the series. The performance of this estimation procedure is illustrated in Section 5.

4. Monte Carlo experiments

In order to assess the performance of the MDEFF estimator, several Monte Carlo experiments are carried out in this section. Two parameter specifications are considered, fractional noise models and ARFIMA(0, d, 1) models. For a fractional noise model, the values of d considered in this study are 0.20, 0.40 and 0.45. The first value corresponds to a *moderate* long memory level and the last two values correspond to *very high* long memory levels. For the ARFIMA(0, d, 1) model, we consider the parameters $d = 0.40$ and $\theta \in \{0.8, 0.4, -0.4, -0.8\}$. The MDEFF method is run in this case with $M = 10$ autocorrelations and $d_0 = 0.5$.

For each simulation run, we generate 1000 replications of Gaussian long-memory time series of sizes $n = 100, 250, 500$. The quantities reported in this simulation study are: the mean of the estimators, the standard deviation and the root mean squared error (RMSE). The MDEFF approach is compared to other three estimators: the approximate MLE proposed by Haslett and Raftery (1989) and implemented in *R*, the Whittle (1953) estimator and the state space estimate proposed by Chan and Palma (1998). These estimators are hereafter referred to as HR, Whittle and Kalman, respectively. The MDEFF, Whittle and Kalman estimates are calculated by means of *R* codes, which are available upon request.

For comparison purposes, Tables 2 and 3 contain the theoretical asymptotic MDEFF standard deviations of $\widehat{d}$ and $\widehat{\theta}$ for $M = 10$ autocorrelations and $d_0 = 0.5$. Note that these standard deviations are valid for Gaussian and non-Gaussian processes. In addition, we provide the asymptotic standard deviations for the corresponding Gaussian maximum likelihood estimates.

As evidenced from the simulation results reported in Table 4, the MDEFF method is very good for fractional noise models in terms of bias, and the standard deviations are very close to their theoretical counterparts shown in Table 2. For ARFIMA(0, d, 1) models, Tables 5–7 reveal that the MDEFF estimates perform well in terms of bias when $\theta > 0$. For the case $\theta < 0$, there is some bias when $n = 500$ and the bias seems to increase for smaller sample sizes. Besides, the sample standard deviations of the MDEFF estimates for $n = 500$ and $n = 250$ are very close to their theoretical counterparts shown in Table 3. This also occurs for the estimate $\widehat{d}$ for a sample size $n = 100$ with positive θ and for $\widehat{\theta}$ with $\theta = 0.4$. For ARFIMA(0, d, 1) models, the worst performance corresponds to $\theta = -0.4$. A similar situation is observed for the standard deviations of the MLE, as evidenced by Table 3.

Next, we compare the performance of the MDEFF method with three other maximum likelihood estimators. For fractional noise models, we note in Table 4 that the MDEFF outperforms the other three methods in terms of bias, particularly when $n = 100$ and $d = 0.45$. The HR estimates display the smallest standard deviations, closely followed by the MDEFF estimates. On the other hand, regarding the RMSE, the MDEFF is the best method, excepting the cases $d = 0.2$ for $n = 250$ and $n = 500$. From Tables 5 to 7, for ARFIMA(0, d, 1) models with positive θ, the MDEFF seems to be the best method in terms of bias of $\widehat{d}$

Table 4
Monte Carlo experiments. 1000 replications of Gaussian ARFIMA(0, d, 0) processes of size n. MDEFF method applied with $d_0 = 0.5$ and $M = 10$ autocorrelations.

n	d	MDEFF	HR	Whittle	Kalman
100	0.45	0.4307[a]	0.3695	0.4302	0.3923
		0.0646[b]	0.0722	0.0848	0.0833
		0.0674[c]	0.1081	0.0870	0.1013
	0.4	0.3906	0.3346	0.3907	0.3521
		0.0795	0.0770	0.0960	0.0862
		0.0800	0.1010	0.0964	0.0986
	0.2	0.2064	0.1499	0.1623	0.1575
		0.0928	0.0853	0.1032	0.0862
		0.0929	0.0988	0.1098	0.0960
250	0.45	0.4401	0.4129	0.4544	0.4315
		0.0460	0.0447	0.0476	0.0520
		0.0471	0.0581	0.0477	0.0551
	0.4	0.3982	0.3725	0.4114	0.3863
		0.0516	0.0478	0.0568	0.0543
		0.0517	0.0551	0.0579	0.0560
	0.2	0.2012	0.1772	0.1869	0.1816
		0.0584	0.0528	0.0590	0.0530
		0.0584	0.0575	0.0604	0.0561
500	0.45	0.4451	0.4297	0.4609	0.4528
		0.0335	0.0313	0.0343	0.0375
		0.0338	0.0373	0.0360	0.0376
	0.4	0.3997	0.3866	0.4127	0.4040
		0.0374	0.0338	0.0388	0.0397
		0.0374	0.0363	0.0408	0.0398
	0.2	0.1995	0.1881	0.1953	0.1925
		0.0418	0.0365	0.0391	0.0374
		0.0418	0.0384	0.0393	0.0382

[a] Is the sample mean for each combination.
[b] Is the standard deviation for each combination.
[c] Is the RMSE for each combination.

Table 5
Monte Carlo experiments. 1000 replications of Gaussian ARFIMA(0, d, 1) processes of size $n = 100$ with $d = 0.4$. MDEFF method applied with $d_0 = 0.5$ and $M = 10$ autocorrelations.

θ	MDEFF		HR		Whittle		Kalman	
	$\widehat{d}$	$\widehat{\theta}$	$\widehat{d}$	$\widehat{\theta}$	$\widehat{d}$	$\widehat{\theta}$	$\widehat{d}$	$\widehat{\theta}$
0.8	0.3659[a]	0.7444	0.3342	0.8172	0.3565	0.7482	0.3456	0.8291
	0.0888[b]	0.1150	0.0810	0.0712	0.1121	0.0922	0.0922	0.0873
	0.0951[c]	0.1277	0.1043	0.0732	0.1202	0.1057	0.1070	0.0920
0.4	0.3714	0.3904	0.3019	0.4613	0.3528	0.4348	0.3323	0.4358
	0.0970	0.1154	0.0908	0.1027	0.1288	0.1138	0.1071	0.1157
	0.1011	0.1158	0.1336	0.1195	0.1371	0.1189	0.1267	0.1211
−0.4	0.2931	−0.2652	0.1693	−0.1813	0.2263	−0.2155	0.2608	−0.2831
	0.1718	0.1987	0.1192	0.1379	0.1829	0.1712	0.1720	0.1970
	0.2023	0.2401	0.2596	0.2585	0.2522	0.2516	0.2212	0.2290
−0.8	0.3178	−0.6846	0.0500	−0.5038	0.2993	−0.7890	0.1195	−0.5612
	0.1926	0.2529	0.0993	0.1645	0.2047	0.2656	0.1755	0.2081
	0.2094	0.2779	0.3638	0.3387	0.2281	0.2657	0.3309	0.3167

[a] Is the sample mean for each combination.
[b] Is the standard deviation for each combination.
[c] Is the RMSE for each combination.

for $n = 100$ and the second best for $n = 250$ and $n = 500$. For $\widehat{\theta}$, the bias of the MDEFF approach is similar to the bias of the Kalman and Whittle methods, excepting the case $\theta = 0.4$ and $n = 100$. In this situation, the bias of the MDEFF approach seems to be considerably smaller than the bias of the Kalman and Whittle methods. Now, regarding the root mean squared error, the MDEFF methodology exhibits small values for $\widehat{d}$ when $n = 100$ and $n = 250$, and for $\widehat{\theta}$ when $n = 100$. On the other hand, the HR estimate of $\widehat{\theta}$ performs well.

For ARFIMA(0, d, 1) models with negative θ, the simulation results show that for $\theta = -0.8$, the HR, Whittle and Kalman methods exhibit severe bias for $\widehat{d}$ and $\widehat{\theta}$, even with a relatively large sample size of $n = 500$. On the contrary, the performance of the MDEFF method measured in terms of bias, is very good. Excepting for $\widehat{\theta}$ when $n = 100$, where the bias is moderate

Table 6
Monte Carlo experiments. 1000 replications of Gaussian ARFIMA(0, d, 1) processes of size $n = 250$ with $d = 0.4$. MDEFF method applied with $d_0 = 0.5$ and $M = 10$ autocorrelations.

θ	MDEFF		HR		Whittle		Kalman	
	$\widehat{d}$	$\widehat{\theta}$	$\widehat{d}$	$\widehat{\theta}$	$\widehat{d}$	$\widehat{\theta}$	$\widehat{d}$	$\widehat{\theta}$
0.8	0.3843[a]	0.7780	0.3737	0.8048	0.3998	0.7697	0.3818	0.8011
	0.0567[b]	0.0625	0.0512	0.0399	0.0637	0.0562	0.0591	0.0474
	0.0588[c]	0.0662	0.0575	0.0401	0.0637	0.0638	0.0618	0.0474
0.4	0.3915	0.3889	0.3593	0.4235	0.4046	0.4014	0.3839	0.4057
	0.0637	0.0765	0.0580	0.0695	0.0726	0.0753	0.0690	0.0757
	0.0643	0.0773	0.0708	0.0733	0.0727	0.0753	0.0708	0.0759
−0.4	0.3518	−0.3392	0.2742	−0.2752	0.3474	−0.3265	0.3567	−0.3630
	0.1352	0.1523	0.0955	0.1117	0.1338	0.1347	0.1288	0.1455
	0.1435	0.1639	0.1579	0.1674	0.1437	0.1534	0.1359	0.1500
−0.8	0.3658	−0.7545	0.1099	−0.5376	0.1285	−0.5476	0.1725	−0.5871
	0.1607	0.1642	0.1300	0.1498	0.1599	0.1599	0.1843	0.1853
	0.1643	0.1703	0.3179	0.3021	0.3151	0.2987	0.2927	0.2822

[a] Is the sample mean for each combination.
[b] Is the standard deviation for each combination.
[c] Is the RMSE for each combination.

Table 7
Monte Carlo experiments. 1000 replications of Gaussian ARFIMA(0, d, 1) time series of size $n = 500$ with $d = 0.4$. MDEFF method applied with $d_0 = 0.5$ and $M = 10$ autocorrelations.

θ	MDEFF		HR		Whittle		Kalman	
	$\widehat{d}$	$\widehat{\theta}$	$\widehat{d}$	$\widehat{\theta}$	$\widehat{d}$	$\widehat{\theta}$	$\widehat{d}$	$\widehat{\theta}$
0.8	0.3903[a]	0.7884	0.3849	0.8030	0.4063	0.7813	0.3982	0.7932
	0.0431[b]	0.0461	0.0365	0.0283	0.0437	0.0364	0.0435	0.0306
	0.0441[c]	0.0475	0.0395	0.0285	0.0441	0.0409	0.0435	0.0314
0.4	0.3930	0.3970	0.3777	0.4138	0.4106	0.3974	0.4045	0.3977
	0.0458	0.0534	0.0399	0.0489	0.0483	0.0526	0.0497	0.0534
	0.0463	0.0535	0.0457	0.0508	0.0495	0.0526	0.0499	0.0534
−0.4	0.3676	−0.3613	0.3236	−0.3221	0.3904	−0.3750	0.4044	−0.4022
	0.1111	0.1231	0.0704	0.0825	0.0939	0.0997	0.0969	0.1080
	0.1157	0.1289	0.1038	0.1134	0.0943	0.1027	0.0969	0.1079
−0.8	0.3820	−0.7806	0.2070	−0.6319	0.2507	−0.6570	0.2532	−0.6559
	0.1276	0.1126	0.1334	0.1372	0.1674	0.1556	0.1838	0.1757
	0.1288	0.1142	0.2345	0.2169	0.2242	0.2113	0.2352	0.2272

[a] Is the sample mean for each combination.
[b] Is the standard deviation for each combination.
[c] Is the RMSE for each combination.

but still smaller than the HR and Kalman methods. In terms of root mean squared error, the MDEFF seems to be the best method. When $\theta = -0.4$, the overall performance of the MDEFF approach is comparable to the best techniques (Whittle and Kalman), but it seems to be better than these two methods for $\widehat{d}$ when $n = 100$.

In summary, according to the Monte Carlo studies, the performance of the MDEFF seems to be very good and comparable to the full likelihood methods. In fact, it seems to be the best methodology for small samples and very negative values of θ. Furthermore, the simulations show that in order to apply the MDEFF method, selecting $M = 10$ autocorrelations seems to be sufficient to produce good results. As reported in Table 1, even though the theoretical efficiency of the MDEFF estimator is relatively low for some negative values of θ, the simulation results indicate that MDEFF outperforms the other methods in terms of root mean squared errors. The computational efficiency of the MDEFF method is an additional relevant advantage over other techniques. This method has a numerical complexity of order $\mathcal{O}(n \log n)$, which is similar to FFT-based methodologies, see for example Chan (1989) and Chen et al. (2006), and it is better than methods based on the Durbin–Levinson algorithm, which is $\mathcal{O}(n^2)$.

5. Illustrations

Three applications of the MDEFF method are presented in this section. The first two illustrations are concerned with the statistical analysis of real-life time series consisting of tree rings and stock trading volume data. The third application shows the performance of the MDEFF in computer-generated long-range dependent data with missing observations.

(a) *Tree rings*. Fig. 1 displays the width of yearly *pinus longaeva* tree rings collected in Mammoth Creek, Utah, USA, for the period 1 AD–1989 AD. This time series is available from the *Time Series Data Library*. Evidence of long-memory behavior can be obtained from the sample autocorrelation function depicted in Fig. 2. This figure also shows that the data distribution appears to be asymmetric.

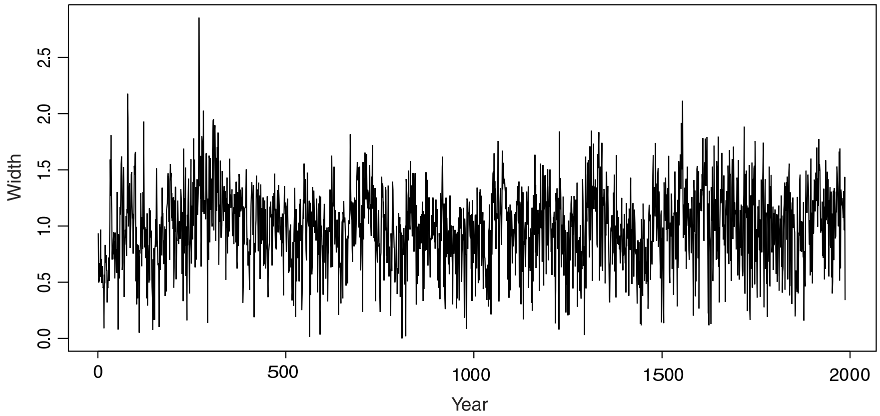

Fig. 1. Tree rings time series.

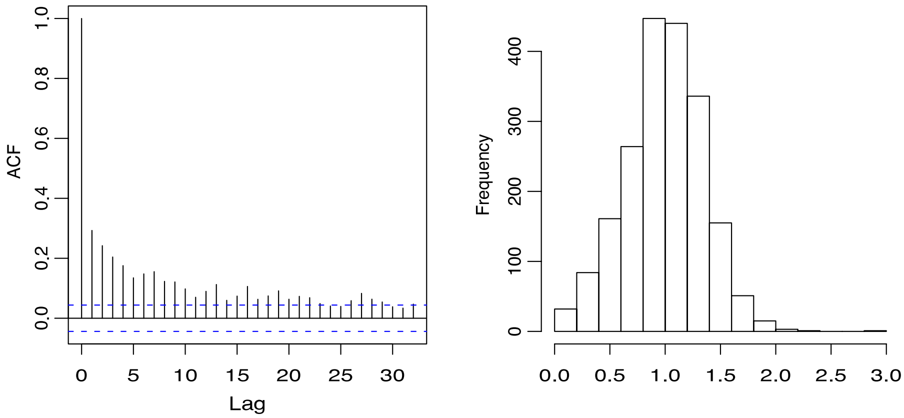

Fig. 2. Autocorrelation function (left) and histogram (right) of the tree rings time series.

Table 8
Tree rings. Estimates of d in the fractional noise model. The standard deviations of the MDEFF estimates based on M autocorrelations are reported within parentheses.

Period	M	MDEFF	HR	Whittle	Kalman
1 AD–989 AD	10	0.232 (0.020)	0.237	0.241	0.247
	20	0.229 (0.019)			
990 AD–1989 AD	10	0.289 (0.029)	0.298	0.309	0.314
	20	0.302 (0.028)			
1 AD–1989 AD	10	0.195 (0.028)	0.187	0.192	0.195
	20	0.184 (0.026)			

Since the tree ring time series seems to exhibit different behavior in terms of memory, we split the series into two periods, 1 AD–989 AD and 990 AD–1989 AD. Subsequently, a fractional noise model was fitted to each period by using the MDEFF method with $d_0 = 0.5$, $M = 10$ and $M = 20$ autocorrelations. The results from this analysis, displayed in Table 8, show that the estimates of the Hurst parameter are highly significant, according to the standard deviations reported within parentheses in the fourth column. The memory of the first period is lower than the memory in the second period, and the memory of the full period lies between them. In addition, using 10 or 20 autocorrelations provides almost the same estimates of the Hurst parameter. This is an important result because if very different estimates were obtained we would have an indication of an inadequate model specification.

Note that given the nature of MDEFF methodology, no explicit assumption on the distribution of the input noise has been made. Besides, by assuming a Gaussian distribution, three other estimates for the fractional noise model are calculated: the approximate maximum likelihood estimates of Haslett–Raftery, and Whittle and the state space Kalman estimator. Table 8 reveals that the four methods provide roughly the same estimates.

(b) *Trading volume*. Long memory in daily stock market volume has been observed by several authors; see for example Lobato and Velasco (2000) and references therein. The dataset analyzed here corresponds to the logarithm of IBM daily

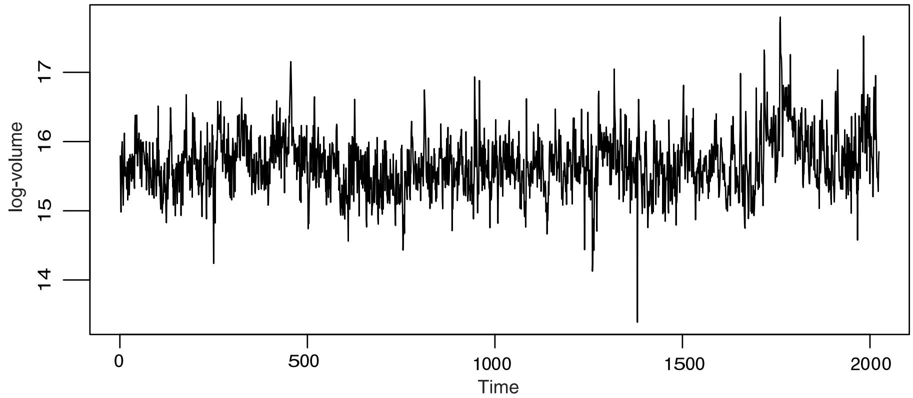

Fig. 3. Trading volume time series.

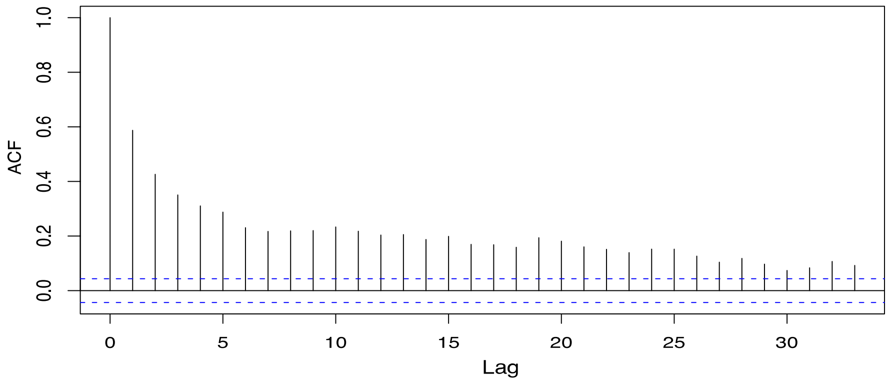

Fig. 4. Autocorrelation function of the trading volume time series.

Table 9
Trading volume. Estimates of d and θ in the ARFIMA$(0, d, 1)$ model. The standard deviations of the MDEFF estimates based on M autocorrelations are reported within parentheses.

Parameter	M	MDEFF	HR	Whittle	Kalman
d	10	0.341 (0.030)	0.340	0.353	0.370
	20	0.347 (0.028)			
θ	10	0.134 (0.035)	0.137	0.129	0.120
	20	0.122 (0.034)			

trading volume for the period January 2, 1986 to December 31, 1993. This time series, displayed in Fig. 3 consists of 2024 observations, with mean 15.690 and standard deviation 0.422.

The sample ACF, depicted in Fig. 4, shows the presence of long memory with high first autocorrelations. The latter feature implies that a short memory component may be needed in the model. For this reason an ARFIMA$(0, d, 1)$ process was fitted using the MDEFF method with $d_0 = 0.5$ for $M = 10$ and $M = 20$ autocorrelations.

From Table 9, it can be observed that all the MDEFF estimates are highly significant and very robust to the value of M, indicating a good model specification. In addition, note that these estimates are close to the corresponding ones obtained by the HR, Whittle and Kalman methods.

(c) *Missing data*. The performance of the MDEFF method in the context of time series with missing values is examined here by means of Monte Carlo experiments. In these simulations, 1000 replications of normally distributed fractional noise processes of lengths $n = 250, 500$ are generated. Three different percentages of missing data are considered: 5%, 10% and 20%. The MDEFF estimates are calculated based on $M = 10$ autocorrelations and $d_0 = 0.5$ following the procedure described in Section 3.2.2. The sample autocorrelations are calculated by using the *acf* function of the statistical package *R*, which allows for the handling of missing values. On the other hand, the HR estimates are calculated by interpolating the missing values with cubic splines. The results from these simulations are reported in Table 10. This table reveals that the performance of the MDEFF method is very good in terms of point estimation. The same occurs for the empirical standard deviations, excepting the case $d = 0.2$ with 20% of missing data. Note that the bias of the HR method increases with the percentage of missing

Table 10
Monte Carlo experiments. 1000 replications of Gaussian ARFIMA(0, d, 0) time series of size n with 5%, 10% and 20% of missing data. MDEFF method applied with $d_0 = 0.5$ and $M = 10$ autocorrelations.

n	d	5%		10%		20%	
		MDEFF	HR	MDEFF	HR	MDEFF	HR
250	0.4	0.3950[a]	0.3914	0.3963	0.4140	0.4013	0.4540
		0.0569[b]	0.0475	0.0596	0.0437	0.0689	0.0315
		0.0571[c]	0.0483	0.0597	0.0459	0.0689	0.0625
	0.2	0.2049	0.2098	0.2019	0.2422	0.2184	0.3224
		0.0677	0.0539	0.0775	0.0527	0.1021	0.0600
		0.0678	0.0547	0.0775	0.0675	0.1037	0.1363
500	0.4	0.3976	0.4064	0.4009	0.4300	0.4038	0.4704
		0.0401	0.0334	0.0445	0.0311	0.0528	0.0196
		0.0402	0.0339	0.0445	0.0432	0.0529	0.0731
	0.2	0.1960	0.2176	0.1937	0.2518	0.2005	0.3353
		0.0474	0.0365	0.0578	0.0386	0.0753	0.0423
		0.0475	0.0406	0.0581	0.0646	0.0753	0.1417

[a] Is the sample mean for each combination.
[b] Is the standard deviation for each combination.
[c] Is the RMSE for each combination.

values. In particular, for 20% of missing values, the HR estimate is severely biased. In addition, the MDEFF method displays less bias compared to the HR approach in all these cases studied.

6. Conclusions

A new estimation method based on minimum distance and a fractional filtering technique has been proposed and compared to three well known methods: the approximate MLE of Haslett and Raftery (1989), a quasi-likelihood method proposed by Whittle (1953) and the state space approach of Chan and Palma (1998). The main features of the MDEFF estimator can be summarized as follows. It is easy to calculate and to implement, since the computation of the MDEFF estimates is based on a reduced number of sample autocorrelations. The MDEFF approach allows for the estimation of Gaussian and non-Gaussian ARFIMA processes with uncorrelated innovations. No explicit assumptions on the distribution of the data have to be made. The estimator of d is not affected by the estimation of the level parameter μ. Several Monte Carlo experiments indicate that the MDEFF method exhibits very small bias and its precision, measured by the mean squared error, is very good when compared to the other techniques considered in this study. In addition, for small sized time series and ARFIMA(0, d, 1) models with negative θ, the performance of the MDEFF approach is substantially better than the performance of the other methods. The MDEFF works well even with missing data. Finally, given its numerical efficiency, the MDEFF technique is useful for handling huge datasets, typically found in the context of long-memory time series.

Acknowledgments

This research was partially supported by CNPq 474329 and FAEPEX grants for the first author and grants Fondecyt 1120758 and ACT 1112 from Conicyt PIA for the second author.

Appendix

This Appendix provides the proofs of Proposition 1 and Theorem 1. First, we state Lemma 1 which is needed to show Proposition 1 and subsequently we present Lemma 2 which is useful for proving Theorem 1.

Lemma 1. *Let $\{y_t\}$ be a stationary process with spectral density $f(\cdot)$ and autocovariance function $\gamma(\cdot)$. If f^2 is integrable, then for any $h \in \mathbb{Z}$,*

$$\sum_{j=-\infty}^{\infty} \gamma(j)\gamma(j+h) = 2\pi \int_{-\pi}^{\pi} f^2(\lambda) e^{-ih\lambda} d\lambda. \tag{14}$$

Proof. Let $a_n = 2\pi \int_{-\pi}^{\pi} f^2(\lambda) e^{-i\lambda h} d\lambda - \sum_{j=-n}^{n} \gamma(j)\gamma(j+h)$. Since f^2 is integrable, $\{a_n\}$ is a well-defined sequence for $n \geq 0$. Furthermore, we can write a_n as

$$a_n = 2\pi \int_{-\pi}^{\pi} [f(\lambda) - f_n(\lambda)] f(\lambda) e^{-i\lambda h} d\lambda,$$

where $f_n(\lambda) = \frac{1}{2\pi}\sum_{j=-n}^{n}\gamma(j)e^{-i\lambda j}$. Note that by Bessel's inequality,

$$(2\pi)^{-2}\sum_{j=0}^{\infty}\gamma^2(j) = \sum_{j=0}^{\infty}|\langle f\ ,\ e^{ij\,\cdot}\rangle|^2 \le \|f\|_2^2 = \int_{-\pi}^{\pi} f(\lambda)^2 d\lambda < \infty.$$

Since $\sum_{j=-\infty}^{\infty}\gamma(j)^2 < \infty$, $\{f_n\}$ is a Cauchy sequence in $\mathcal{L}_2(d\lambda)$ converging to the limit

$$f(\lambda) = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\gamma(j)e^{-i\lambda j},$$

in $\mathcal{L}_2(d\lambda)$. Thus, $\|f - f_n\|_2 \to 0$ as $n \to \infty$. Consequently, by the Cauchy–Schwarz inequality $|a_n| \le 2\pi\int_{-\pi}^{\pi}|f(\lambda) - f_n(\lambda)|f(\lambda)d\lambda \le 2\pi\|f - f_n\|_2\|f\|_2$. Therefore, since $\|f\|_2 < \infty$ we conclude that $a_n \to 0$ as $n \to \infty$ and the result follows. □

Lemma 2. *Under the assumptions of Theorem 1, there is a positive constant K such that*

$$\|x_t - \tilde{x}_t\| \le K\ t^{d-d_0-\frac{1}{2}},$$

for all $t \ge 1$.

Proof. Observe that $x_t - \tilde{x}_t = \sum_{j=t}^{\infty}\pi_j(d_0)y_{t-j}$. Therefore,

$$\begin{aligned}\|x_t - \tilde{x}_t\|^2 &= \text{Var}\left(\sum_{j=t}^{\infty}\pi_j(d_0)y_{t-j}\right) = \sum_{i=t}^{\infty}\sum_{j=t}^{\infty}\pi_i(d_0)\pi_j(d_0)\gamma(i-j)\\ &= \gamma(0)\sum_{j=t}^{\infty}\pi_j^2(d_0) + 2\sum_{i>j\ge t}^{\infty}\pi_i(d_0)\pi_j(d_0)\gamma(i-j)\\ &= \gamma(0)\sum_{j=t}^{\infty}\pi_j^2(d_0) + 2\sum_{j=t}^{\infty}\sum_{k=1}^{\infty}\pi_j(d_0)\pi_{j+k}(d_0)\gamma(k)\\ &= \gamma(0)\sum_{j=t}^{\infty}\pi_j^2(d_0) + 2\sum_{j=t}^{\infty}\pi_j(d_0)\left[\sum_{k=1}^{\infty}\pi_{j+k}(d_0)\gamma(k)\right].\end{aligned}$$

By Lemma 3.2 of Palma (2007) we have that

$$\sum_{j=t}^{\infty}\pi_j^2(d_0) \sim C_1\ t^{-2d_0-1}, \qquad \sum_{k=1}^{\infty}\pi_{j+k}(d_0)\gamma(k) \sim C_2\ j^{2d-d_0-1}.$$

Consequently, since $\pi_j(d_0) \sim j^{d_0-1}$,

$$\sum_{j=t}^{\infty}\pi_j(d_0)\left[\sum_{k=1}^{\infty}\pi_{j+k}(d_0)\gamma(k)\right] \sim C_2\sum_{j=t}^{\infty}\pi_j(d_0)j^{2d-d_0-1} \sim C_3\ t^{2d-2d_0-1}.$$

Therefore, $\|x_t - \tilde{x}_t\| \le\ K\ t^{d-d_0-\frac{1}{2}}$, as required. □

Remark 2. If $f(\lambda)$ is the spectral density of an ARFIMA(p, d, q) process

$$\phi(B)y_t = \theta(B)(1 - B)^{-d}\varepsilon_t, \tag{15}$$

then $f^2(\lambda)$ can be written as $f^2(\lambda) = \frac{\sigma^2}{2\pi}g(\lambda)$ where $g(\lambda)$ is the spectral density of the ARFIMA(2p, 2d, 2q) process

$$\phi(B)^2y_t = \theta(B)^2(1 - B)^{-2d}\varepsilon_t. \tag{16}$$

If $\{\rho_1, \ldots, \rho_p\}$ are the simple roots of the AR part of the process (15) then the roots of the AR part of process (16) are also $\{\rho_1, \ldots, \rho_p\}$ but they have multiplicity 2. The spectral density of an ARFIMA(2p, 2d, 2q) process is not integrable for $d > \frac{1}{4}$.

Proof of Proposition 1. Let $\omega = e^{-i\lambda}$ and let $g(\lambda)$ be the spectral density of the ARFIMA(2p, 2d, 2q) process (16) with $d < \frac{1}{4}$. Then $g(\lambda)$ is given by

$$g(\lambda) = \frac{\sigma^2}{2\pi}|1 - \omega|^{-4d}|\theta(\omega)|^4|\phi(\omega)|^{-4}$$

$$= \frac{\sigma^2}{2\pi}|1-\omega|^{-4d}|\theta(\omega)|^4 \prod_{j=1}^{p}(1-\rho_j\omega)^{-2}(1-\rho_j\omega^{-1})^{-2}$$

$$= \frac{\sigma^2}{2\pi}|1-\omega|^{-4d} \sum_{k=-2q}^{2q}\varphi(k)\omega^k \prod_{j=1}^{p}(1-\rho_j\omega)^{-2}(1-\rho_j\omega^{-1})^{-2}.$$

By partial fraction decomposition g can be written as

$$g(\lambda) = \frac{\sigma^2}{2\pi}|1-\omega|^{-4d} \sum_{k=-2q}^{2q}\varphi(k)\sum_{j=1}^{p}c_j\left\{\frac{\alpha_j}{1-\rho_j\omega} + \frac{\beta_j}{1-\rho_j^{-1}\omega} + \frac{\rho_j^{4p-2}}{(1-\rho_j\omega)^2} + \frac{\rho_j^{-2}}{(1-\rho_j^{-1}\omega)^2}\right\}\omega^{2p+k}. \tag{17}$$

Let,

$$D(d) = \frac{\Gamma(1-4d)}{\Gamma(1-2d)\Gamma(2d)}\frac{\Gamma(h+2d)}{\Gamma(1+h-2d)},$$

as in Sowell (1992, p. 184ff), for any $h \in \mathbb{Z}$ we have:

$$\begin{aligned}\int_{-\pi}^{\pi}\frac{|1-\omega|^{-4d}}{1-\rho\omega}\,\omega^h\,d\lambda &= \sum_{j=0}^{\infty}\rho^j\int_{-\pi}^{\pi}|1-\omega|^{-4d}\,\omega^{h+j}\,d\lambda\\ &= 2\pi D(d)F(2d+h,1;1+h-2d;\rho).\end{aligned}$$

By analogous calculations we get

$$\begin{aligned}\int_{-\pi}^{\pi}\frac{|1-\omega|^{-4d}}{1-\rho^{-1}\omega}\,\omega^h\,d\lambda &= \int_{-\pi}^{\pi}\left[1-\sum_{j=0}^{\infty}\rho^j\omega^{-j}\right]|1-\omega|^{-4d}\,\omega^h\,d\lambda\\ &= 2\pi D(d)[1-F(2d-h,1;1-h-2d;\rho)],\\ \int_{-\pi}^{\pi}\frac{|1-\omega|^{-4d}}{(1-\rho\omega)^2}\,\omega^h\,d\lambda &= \sum_{j=0}^{\infty}(j+1)\rho^j\int_{-\pi}^{\pi}|1-\omega|^{-4d}\,\omega^{h+j}\,d\lambda\\ &= 2\pi D(d)F(h+2d,2;1+h-2d;\rho),\\ \int_{-\pi}^{\pi}\frac{|1-\omega|^{-4d}}{(1-\rho^{-1}\omega)^2}\,\omega^h\,d\lambda &= \sum_{j=2}^{\infty}(j-1)\rho^j\int_{-\pi}^{\pi}|1-\omega|^{-4d}\,\omega^{h-j}\,d\lambda\\ &= 2\pi D(d)[F(2d-h,2;1-h-2d;\rho)-2F(2d-h,1;1-h-2d;\rho)+1].\end{aligned}$$

Now, by Lemma 1 and Remark 2,

$$\sum_{j=-\infty}^{\infty}\gamma(j)\gamma(j+s) = 2\pi\int_{-\pi}^{\pi}f^2(\lambda)e^{-is\lambda}d\lambda = \sigma^2\int_{-\pi}^{\pi}g(\lambda)\omega^s d\lambda,$$

then by combining the above expressions with the spectral density decomposition (17) we get the required result. □

Proof of Theorem 1. Part (a). Since the process $\{x_t\}$ has long-memory parameter $d-d_0$ with $-1 < d-d_0 < \frac{1}{4}$, this process is second order stationary and invertible; cf. Bondon and Palma (2007) and Theorem 3.4 of Palma (2007). Observe that the spectral density of $\{x_t\}$ satisfies $f(\lambda) \sim c|\lambda|^{2(d_0-d)}$ as $|\lambda| \to 0$, for some positive constant c. Thus, $\int_{-\pi}^{\pi} f(\lambda)^2\,d\lambda \le K\int_{-\pi}^{\pi}\lambda^{4(d_0-d)}\,d\lambda$ for some positive constant K. Consequently, $f(\lambda)^2$ is integrable for $-\infty < d-d_0 < \frac{1}{4}$. Therefore, an application of the central limit theorem of Hannan (1976) yields $\sqrt{n}\,(\widehat{\rho}_x-\rho_x) \to N(0,V)$ in distribution, as $n\to\infty$ for $-\infty < d-d_0 < \frac{1}{4}$. Thus, it suffices to show that $\sqrt{n}(\widehat{\rho}_x-\widehat{\rho}_{\tilde{x}}) \to 0$ in probability as $n\to\infty$. To this end, note that

$$\widehat{\rho}_x(k)-\widehat{\rho}_{\tilde{x}}(k) = \frac{\widehat{\gamma}_x(k)-\widehat{\gamma}_{\tilde{x}}(k)}{\widehat{\gamma}_x(0)} + \widehat{\gamma}_{\tilde{x}}(k)\frac{\widehat{\gamma}_{\tilde{x}}(0)-\widehat{\gamma}_x(0)}{\widehat{\gamma}_{\tilde{x}}(0)\widehat{\gamma}_x(0)}.$$

By the ergodicity of $\{x_t\}$, for fixed k we have that $\widehat{\gamma}_x(k) \to \gamma_x(k)$ in probability, as $n\to\infty$ and then as a consequence of the result (18), $\widehat{\gamma}_{\tilde{x}}(k)$ converges to a constant in probability. Therefore, by Slutsky's theorem it suffices to prove that $\sqrt{n}[\widehat{\gamma}_x(k)-\widehat{\gamma}_{\tilde{x}}(k)] \to 0$ in probability as $n\to\infty$, for a fixed k. But, this term can be written as

$$\sqrt{n}[\widehat{\gamma}_x(k)-\widehat{\gamma}_{\tilde{x}}(k)] = \frac{1}{\sqrt{n}}\sum_{t=1}^{n-k}\left[(x_t-\tilde{x}_t+\tilde{x}-\bar{x})(\tilde{x}_{t+k}-\tilde{x})+(x_{t+k}-\tilde{x}_{t+k}+\tilde{x}-\bar{x})(x_t-\bar{x})\right],$$

where $\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t$ and $\tilde{x} = \frac{1}{n}\sum_{t=1}^{n} \tilde{x}_t$. Therefore,

$$\sqrt{n}E|\widehat{\gamma}_x(k) - \widehat{\gamma}_{\tilde{x}}(k)| \leq \frac{1}{\sqrt{n}} \sum_{t=1}^{n-k} \left[E|(x_t - \tilde{x}_t + \tilde{x} - \bar{x})(\tilde{x}_{t+k} - \tilde{x})| + E|(x_{t+k} - \tilde{x}_{t+k} + \tilde{x} - \bar{x})(x_t - \bar{x})| \right].$$

Thus, an application of the Cauchy–Schwarz inequality yields

$$\begin{aligned} \sqrt{n}E|\widehat{\gamma}_x(k) - \widehat{\gamma}_{\tilde{x}}(k)| &\leq \frac{1}{\sqrt{n}} \sum_{t=1}^{n-k} \left[\|x_t - \tilde{x}_t + \tilde{x} - \bar{x}\| \, \|\tilde{x}_{t+k} - \tilde{x}\| + \|x_{t+k} - \tilde{x}_{t+k} + \tilde{x} - \bar{x}\| \, \|x_t - \bar{x}\| \right] \\ &\leq \frac{1}{\sqrt{n}} \sum_{t=1}^{n-k} \left[(\|x_t - \tilde{x}_t\| + \|\tilde{x} - \bar{x}\|)\|\tilde{x}_{t+k} - \tilde{x}\| + (\|x_{t+k} - \tilde{x}_{t+k}\| + \|\tilde{x} - \bar{x}\|)\|x_t - \bar{x}\| \right]. \end{aligned}$$

Note that by Lemma 2 we have that for all $t \geq 1$, $\|x_t - \tilde{x}_t\| \leq K\ t^{d-d_0-\frac{1}{2}}$. Consequently,

$$\|\tilde{x} - \bar{x}\| = \left\| \frac{1}{n} \sum_{t=1}^{n} (\tilde{x}_t - x_t) \right\| \leq \frac{1}{n} \sum_{t=1}^{n} \|\tilde{x}_t - x_t\| \leq K\ n^{d-d_0-\frac{1}{2}}.$$

Thus, $\sqrt{n}E|\widehat{\gamma}_x(k) - \widehat{\gamma}_{\tilde{x}}(k)| \leq K\ n^{d-d_0}$. Now, since $d < d_0$, we conclude that

$$\sqrt{n}E|\widehat{\gamma}_x(k) - \widehat{\gamma}_{\tilde{x}}(k)| \to 0, \tag{18}$$

as $n \to \infty$. Then, by virtue of Markov's inequality we obtain the result. Part (b). This result is a consequence of Part (a) above and a direct application of Appendix A of Baillie and Chung (2001). □

References

Baillie, R., Chung, H., 2001. Estimation of GARCH models from the autocorrelations of the squares of a process. Journal of Time Series Analysis 22, 631–650.
Beran, J., 1994. Statistics for Long-Memory Processes. Chapman and Hall, New York.
Bhardwaj, G., Swanson, N.R., 2006. An empirical investigation of the usefulness of ARFIMA models for predicting macroeconomic and financial time series. Journal of Econometrics 131, 539–578.
Bondon, P., Palma, W., 2007. A class of antipersistent processes. Journal of Time Series Analysis 28, 261–273.
Chan, R.H., 1989. Circulant preconditioners for Hermitian Toeplitz systems. SIAM Journal on Matrix Analysis and Applications 10, 542–550.
Chan, N.H., Palma, W., 1998. State-space modeling of long-memory processes. Annals of Statistics 26, 719–740.
Cheang, W.K., Reinsel, G.C., 2003. Finite sample properties of ML and REML estimators in time series regression models with long memory noise. Journal of Statistical Computation and Simulation 73, 233–259.
Chen, W., Hurvich, C.M., Lu, Y., 2006. On the correlation matrix of the discrete Fourier transform and the fast solution of large Toeplitz systems for long-memory time series. Journal of the American Statistical Association 101, 812–822.
Cheung, Y.W., Diebold, F.X., 1994. On maximum likelihood estimation of the differencing parameter of fractionally-integrated noise with unknown mean. Journal of Econometrics 62, 317–350.
Doornik, J.A., Ooms, M., 2003. Computational aspects of maximum likelihood estimation of autoregressive fractionally integrated moving average models. Computational Statistics and Data Analysis 42, 333–348.
Doukhan, P., Oppenheim, G., Taqqu, M.S. (Eds.), 2003. Theory and Applications of Long-Range Dependence. Birkhäuser.
Granger, C.W.J., Joyeux, R., 1980. An introduction to long-memory time series models and fractional differencing. Journal of Time Series Analysis 1, 15–29.
Hannan, E.J., 1976. The asymptotic distribution of serial covariances. Annals of Statistics 4, 396–399.
Haslett, J., Raftery, A.E., 1989. Space–time modelling with long-memory dependence: assessing Ireland's wind power resource (with discussion). Applied Statistics 38, 1–21.
Henry, M., 2001. Averaged periodogram spectral estimation with long-memory conditional heteroscedasticity. Journal of Time Series Analysis 22, 431–459.
Hosking, J.R.M., 1981. Fractional differencing. Biometrika 68, 165–176.
Hosking, J.R.M., 1996. Asymptotic distributions of the sample mean, autocovariance and autocorrelation of long-memory time series. Journal of Econometrics 73, 261–284.
Kouamé, E., Hili, O., 2008. Minimum distance estimation of k-factors GARMA processes. Statistics and Probability Letters 78, 3254–3261.
Liseo, B., Marinucci, D., Petrella, L., 2001. Bayesian semiparametric inference on long-range dependence. Biometrika 88, 1089–1104.
Lobato, I., Velasco, C., 2000. Long-memory in stock-market trading volume. Journal of Business & Economic Statistics 18, 410–427.
Mayoral, L., 2007. Minimum distance estimation of stationary and non-stationary ARFIMA processes. Econometrics Journal 10, 124–148.
Palma, W., 2007. Long-Memory Time Series: Theory and Methods. Wiley, New Jersey.
Parzen, E., 1963. On spectral analysis with missing observations and amplitude modulation. Sankhyā. The Indian Journal of Statistics, Series A 25, 383–392.
Robinson, P.M., 1995. Gaussian semiparametric estimation of long range dependence. Annals of Statistics 23, 1630–1661.
Sowell, F., 1992. Maximum likelihood estimation of stationary univariate fractionally integrated time series models. Journal of Econometrics 53, 165–188.
Tieslau, M.A., Schmidt, P., Baillie, R.T., 1996. A minimum distance estimator for long-memory process. Journal of Econometrics 71, 249–264.
Whittle, P., 1953. Estimation and information in stationary time series. Arkiv för Matematik 2, 423–434.