ESTIMATING AND FORECASTING THE LONG MEMORY PARAMETER IN THE PRESENCE OF PERIODICITY

Bisognin, C. and **Lopes, S.R.C.**[1]

Instituto de Matemática - UFRGS, Porto Alegre, RS, Brazil

October 10, 2005

ABSTRACT

We consider one parametric and five semiparametric approaches to estimate D in SARFIMA$(0, D, 0)_s$ processes, that is, when the process is a fractionally integrated ARMA model with seasonality s. We also consider the h-step ahead forecasting for these processes. We present the proof of some features of this model and also a study based on a Monte Carlo simulation for different sample sizes and different seasonal periods. We compare the different estimation procedures analyzing the bias, the mean squared error values, and the confidence intervals for the estimators. We also consider three different methods to choose the total number of regressors in the regression analysis for the semiparametric class of estimation procedures. We apply the methodology to the Nile River flows monthly data, and also to a simulated seasonal fractionally integrated time series.

Keywords: Long Memory Models, Seasonality, Estimation, Forecasting, Bandwidth Size.

1 INTRODUCTION

The ARFIMA(p, d, q) process was first introduced by Granger and Joyeux (1980), and Hosking (1981 and 1984). The most useful feature of this process is the long memory characteristic which is reflected by the hyperbolic decay of its autocorrelation function or by the unboundedness of its spectral density function. While in the ARMA model, dependency between observations decays at a geometric rate.

We want to consider long memory processes with periodicity, and they can be modelled by using the so-called SARFIMA$(p, d, q) \times (P, D, Q)_s$ processes.

Several estimation procedures for the fractional ARFIMA parameter have been proposed, mainly, in the semiparametric and parametric classes. In the first class the regression method proposed by Geweke and Porter-Hudak (1983) was the pioneer. This approach was very important giving rise to several other works. The authors presented a proof when $d \in (-0.5, 0.0)$, nonetheless, the method denoted here by GPH, has been used for a wide range of d. Reisen (1994) proposed a modified form of the regression method based on a smoothed version of the periodogram function.

In the parametric class, the reader will find the methods based on the maximum likelihood function as suggested in Fox and Taqqu (1986), and Sowell (1992), among others. We should point out that the parametric estimator pursued here is the approximated maximum likelihood method proposed by Whittle (1953), and also considered by Fox and Taqqu (1986).

[1]Corresponding author. E-mail: slopes@mat.ufrgs.br

The papers by Porter-Hudak (1990), Ray (1993), Ooms (1995), and Montanari et al. (2000) deal with seasonality analysis for observable data in different fields of application. The work by Hassler (1994) presents a complete generalization of fractional differencing processes with the presence of seasonality treating rigid, and flexible models. It also illustrates the risk of fractional misspecification. The paper by Peiris and Singh (1996) deal with prediction, and minimum mean squared error predictors of one step ahead for seasonal fractionally integrated models. The paper by Reisen and Lopes (1999) present forecasting results for the ARFIMA$(2, d, 2)$ model including the variance of the mean squared error value using the smoothed periodogram regression method for estimating the parameter d. The later paper also presents an analysis of a real observed data comparing the performance of both ARIMA, and ARFIMA models. Ray (1993) forecasts the IBM product revenues using a complete SARFIMA$(p, d, q) \times (P, D, Q)_s$ process.

The main goal of this paper is to analyze the estimation procedures based on five semiparametric and one parametric approaches to estimate D, the *seasonal fractional parameter*, when the model is described by a SARFIMA$(0, D, 0)_s$ process. We prove the ergodicity, some features of the SARFIMA$(0, D, 0)_s$ process, and present the h-step ahead forecasting for it. We apply the methodology to an observed data, and also to a simulated time series.

The paper is organized as follows: Section 2 gives some definitions, and some properties of the SARFIMA$(p, d, q) \times (P, D, Q)_s$ processes. In Section 3 the six estimators of D used here, and based on semiparametric and parametric classes are outlined. Section 4 gives the h-step ahead for the SARFIMA processes, including the mean squared error of forecasting, the forecasting error, and the confidence interval for the forecasting values. Monte Carlo simulations results are in Section 5. An application of the model to the Nile River monthly flows at Aswan, and a complete simulated time series analysis are shown in Section 6. Section 7 concludes.

2 SARFIMA$(p, d, q) \times (P, D, Q)_s$ PROCESSES

In practical situations many time series exhibit a periodic pattern. These time series are very common in meteorology, economics, hydrology, and astronomy. Sometimes, even in these fields, the period of the seasonality can depend on time, that is, the autocorrelation structure of the data varies from season to season. Here, in our analysis, we consider the seasonality period constant over seasons.

We shall consider the *seasonal autoregressive fractionally integrated moving average* processes, denoted here by SARFIMA$(p, d, q) \times (P, D, Q)_s$, which are an extension of the ARFIMA (p, d, q) models, proposed by Granger and Joyeux (1980), and Hosking (1981).

The following sub-section gives some definitions, and some properties for the SARFIMA $(p, d, q) \times (P, D, Q)_s$ processes.

2.1 Some Definitions and Properties

Definition 2.1: Let $\{X_t\}_{t\in\mathbb{Z}}$ be a stochastic process given by the expression

$$\phi(\mathcal{B})\Phi(\mathcal{B}^s)\nabla^d\nabla_s^D(X_t - \mu) = \theta(\mathcal{B})\Theta(\mathcal{B}^s)\varepsilon_t, \tag{2.1}$$

where μ is the *mean* of the process, $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a white noise process, s is the seasonal period, $\mathcal{B}$ is the *backward-shift operator*, that is, $\mathcal{B}^k X_t = X_{t-k}$, and $\mathcal{B}^{sk} X_t = X_{t-sk}$, ∇^d, and ∇_s^D are, respectively, the *difference and the seasonal difference operators*, that is, $\nabla_s^D := (1-\mathcal{B}^s)^D = \sum_{k\geqslant 0}\binom{D}{k}(-\mathcal{B}^s)^k$,where $\binom{D}{k} = \Gamma(1+D)/[\Gamma(1+k)\Gamma(1+D-k)]$, and when $s = 1$ and $D = d$ we have $\nabla_s^D = \nabla^d$. The polynomials $\phi(\cdot)$, $\theta(\cdot)$, $\Phi(\cdot)$, and $\Theta(\cdot)$ with degrees p, q, P, and Q, respectively, are defined by

$$\begin{aligned}
\phi(\mathcal{B}) &= \sum_{i=0}^{p} (-\phi_i)\, \mathcal{B}^i, \quad \theta(\mathcal{B}) = \sum_{j=0}^{q} (-\theta_j)\, \mathcal{B}^j, \\
\Phi(\mathcal{B}) &= \sum_{k=0}^{P} (-\Phi_k)\, \mathcal{B}^k, \quad \Theta(\mathcal{B}) = \sum_{l=0}^{Q} (-\Theta_l)\, \mathcal{B}^l,
\end{aligned}$$

where ϕ_i, $1 \leqslant i \leqslant p$, θ_j, $1 \leqslant j \leqslant q$, Φ_k, $1 \leqslant k \leqslant P$, and Θ_l, $1 \leqslant l \leqslant Q$ are constants. Then, $\{X_t\}_{t \in \mathbb{Z}}$ is a *seasonal fractionally integrated* ARMA process with period s, denoted by SARFIMA$(p, d, q) \times (P, D, Q)_s$, where d and D are, respectively, *the degree of differencing and of seasonal differencing parameters.*

Remarks: (1) A particular case of the SARFIMA$(p, d, q) \times (P, D, Q)_s$ process is when $p = q = P = Q = 0$. This process is called the *pure seasonal fractionally integrated model with period s*, denoted by SARFIMA$(0, D, 0)_s$, which will be the mainly goal of our study in this work and it is given by

$$\nabla_s^D (X_t - \mu) \equiv (1 - \mathcal{B}^s)^D (X_t - \mu) = \varepsilon_t, \quad t \in \mathbb{Z}. \tag{2.2}$$

(2) When $P = Q = 0$, $D = 0$ and $s = 1$ the SARFIMA$(p, d, q) \times (P, D, Q)_s$ process is just the ARFIMA(p, d, q) process (see Beran, 1994). In this case we already know the behaviour of the parameter estimators (see Reisen and Lopes, 1999 and Lopes et al., 2004).

We now shall give some of the properties of the SARFIMA$(0, D, 0)_s$ process.

Theorem 2.1: *Let $\{X_t\}_{t \in \mathbb{Z}}$ be the SARFIMA$(0, D, 0)_s$ process given by the expression (2.2), with zero mean and $s \in \mathbb{N}$ as the seasonal period. Then,*

(i) *when $D > -0.5$, $\{X_t\}_{t \in \mathbb{Z}}$ is an invertible process with infinite autoregressive representation given by*

$$\Pi(\mathcal{B}^s) X_t = \sum_{k \geqslant 0} \pi_k X_{t-sk} = \varepsilon_t,$$

where

$$\pi_k = \frac{-D(1 - D) \cdots (k - D - 1)}{k!} = \frac{(k - D - 1)!}{k!(-D - 1)!} = \frac{\Gamma(k - D)}{\Gamma(k + 1)\Gamma(-D)}. \tag{2.3}$$

When $k \to \infty$, $\pi_k \sim \frac{k^{-D-1}}{\Gamma(-D)}$.

(ii) *when $D < 0.5$, $\{X_t\}_{t \in \mathbb{Z}}$ is a stationary process with an infinite moving average representation given by*

$$X_t = \Psi(\mathcal{B}^s) \varepsilon_t = \sum_{k \geqslant 0} \psi_k \varepsilon_{t-sk},$$

where

$$\psi_k = \frac{D(1 + D) \cdots (k + D - 1)}{k!} = \frac{(k + D - 1)!}{k!(D - 1)!} = \frac{\Gamma(k + D)}{\Gamma(k + 1)\Gamma(D)}. \tag{2.4}$$

When $k \to \infty$, $\psi_k \sim \frac{k^{D-1}}{\Gamma(D)}$.

In the following, we assume that $D \in (-0.5, 0.5)$.

(iii) *The process* $\{X_t\}_{t\in\mathbb{Z}}$ *has spectral density function given by*

$$f_X(w) = \frac{\sigma_\varepsilon^2}{2\pi}\left[2\sin\left(\frac{sw}{2}\right)\right]^{-2D}, \quad 0 < w \leqslant \pi. \tag{2.5}$$

At the seasonal frequencies, for $\nu = 0, 1, \cdots, [s/2]$, *where* $[x]$ *means the integer part of* x, *it behaves as*

$$f_X\left(\frac{2\pi\nu}{s} + w\right) \sim f_\varepsilon\left(\frac{2\pi\nu}{s}\right)(sw)^{-2D}, \quad \text{when} \quad w \to 0\cdot$$

In the following, let A *be the set* $\{1, 2, \cdots, s-1\}$, *and* $\mathbb{Z}_\geqslant$ *be the set* $\{k \in \mathbb{Z} | k \geqslant 0\}$.

(iv) *The process* $\{X_t\}_{t\in\mathbb{Z}}$ *has autocovariance and autocorrelation functions of order* k, $k \in \mathbb{Z}_\geqslant$, *given, respectively, by*

$$\gamma_X(sk+\xi) = \begin{cases} \frac{(-1)^k\Gamma(1-2D)}{\Gamma(1+k-D)\Gamma(1-k-D)}\sigma_\varepsilon^2 = \gamma_X(k), & \text{if } \xi = 0 \\ 0, & \text{if } \xi \in A, \end{cases} \tag{2.6}$$

and

$$\rho_X(sk+\xi) = \begin{cases} \frac{\Gamma(k+D)\Gamma(1-D)}{\Gamma(1+k-D)\Gamma(D)} = \rho_X(k), & \text{if } \xi = 0 \\ 0, & \text{if } \xi \in A. \end{cases} \tag{2.7}$$

When $k \to \infty$, $\rho_X(sk) \sim \frac{\Gamma(1-D)}{\Gamma(D)} k^{2D-1}$.

(v) *The process* $\{X_t\}_{t\in\mathbb{Z}}$ *has partial autocorrelation function given by*

$$\phi_X(sk+\xi, sl+\eta) = \begin{cases} -\binom{k}{l}\frac{\Gamma(l-D)\Gamma(k-l+1-D)}{\Gamma(-D)\Gamma(1+k-D)} = \phi_X(k,l), & \text{if } \eta = 0 \\ 0, & \text{if } \eta \in A, \end{cases} \tag{2.8}$$

for any $k, l \in \mathbb{Z}_\geqslant$, *and* $\xi \in A \cup \{0\}$.

From expression (2.8), when $l = k$, *the partial autocorrelation function of order* k *is given by*

$$\phi_X(sk, sk) = \frac{D}{k-D} = \phi_X(k,k), \quad \text{for all} \quad k \in \mathbb{Z}_\geqslant. \tag{2.9}$$

The proof of this theorem can be found in Brietzke *et al.* (2005).

Remarks: (1) The spectral density function of the SARFIMA$(0, D, 0)_s$ process in the seasonal frequencies is unbounded when $0 < D < 0.5$, and it has zeroes when D is negative.

(2) Among seasonal frequencies the SARFIMA processes have similar behaviour as the ARFIMA processes.

(3) The SARFIMA$(p, d, q) \times (P, D, Q)_s$ process is stationary when d, and D are less than 0.5, and the polynomials $\phi(\mathcal{B}) \cdot \Phi(\mathcal{B}) = 0$, and $\theta(\mathcal{B}) \cdot \Theta(\mathcal{B}) = 0$ have no roots in common, and all roots are outside of the unit circle. When $D > 0$, the process has seasonal long memory.

(4) The paper by Brietzke et al. (2005) gives a closed formula for the Durbin-Levinson's algorithm relating the partial autocorrelation and the autocorrelation functions for the seasonal fractionally integrated processes. This algorithm is very important for generating these processes when one uses the method proposed by Hosking (1984) (see Section 5).

(5) The ergodicity of the stochastic process $\{X_t\}_{t\in\mathbb{Z}}$, given by the expression (2.2), with seasonality s and $D < 0.5$ is based on the infinite moving average representation of this process and it is sufficient to show that the coefficients of this representation is square summable. For a complete proof see Bisognin and Lopes (2005).

The next theorem shows that for SARFIMA$(0, D, 0)_s$ processes the conditional expectation and the conditional variance depend only on the past values distant from multiples of the seasonality s. This theorem is very important when one needs to generate the mentioned processes (see Section 5 of this work).

Theorem 2.2: *Let $\{X_t\}_{t\in\mathbb{Z}}$ be a SARFIMA$(0, D, 0)_s$ process given by the expression (2.2), with zero mean, $s \in \mathbb{N}$ as the seasonal period, and $D \in (-0.5, 0.5)$. The conditional expectation and the conditional variance of X_t, given X_l, for all $l < t$, denoted respectively by $m_t \equiv \mathbb{E}(X_t|X_l, l < t)$ and $v_t \equiv Var(X_t|X_l, l < t)$, are given by*

$$\begin{cases} m_\zeta = 0, \quad \textit{for} \quad \zeta \in A, \\ m_{sk} = \displaystyle\sum_{j=1}^{k} \phi_X(sk, sj)\, X_{sk-sj}, \quad \textit{for} \quad k \in \mathbb{N}, \\ m_{sk+\zeta} = \displaystyle\sum_{j=1}^{k} \phi_X(sk+\zeta, sj)\, X_{sk+\zeta-sj}, \quad \textit{for} \quad \zeta \in A, \end{cases} \tag{2.10}$$

and

$$\begin{cases} v_\zeta = \sigma_\varepsilon^2, \quad \textit{for} \quad \zeta \in A, \\ v_{sk} = \sigma_\varepsilon^2 \displaystyle\prod_{j=1}^{k} [1 - \phi_X^2(sj, sj)], \quad \textit{for} \quad k \in \mathbb{N}, \\ v_{sk+\zeta} = v_{sk}, \quad \textit{for} \quad \zeta \in A, \end{cases} \tag{2.11}$$

where $t = \zeta$ determines the mean and the variance for lags smaller than s, $t = sk$ for lags multiple of s, and $t = sk+\zeta$ for lags not multiple of s, $\phi_X(\cdot,\cdot)$ is the partial autocorrelation function of the process $\{X_t\}_{t\in\mathbb{Z}}$ given by item (v) in Theorem 2.1, σ_ε^2 is the variance of the white noise process and A is the set $\{1, 2, \cdots, s-1\}$.

Proof: Let $\{X_t\}_{t\in\mathbb{Z}}$ be a SARFIMA$(0, D, 0)_s$ process with seasonality s, given by the expression (2.2). First, we want to obtain the conditional expectation of the $\{X_t\}_{t\in\mathbb{Z}}$ processes. From Hosking (1984), for any stationary stochastic process with zero mean, of the form (2.2), its conditional expectation, given the past observations, can be written as

$$m_t \equiv \mathbb{E}(X_t|X_\ell, \ell < t) = \sum_{j=1}^{t} \phi_X(t, j) X_{t-j}, \tag{2.12}$$

where $\phi_X(\cdot,\cdot)$ is the partial autocorrelation function of a SARFIMA$(0, D, 0)_s$ process given in item (v) of Theorem 2.1. Let s be in $\mathbb{N}$ and $\zeta \in A$. From expression (2.12), when $t = \zeta$ we have

$$m_\zeta = \sum_{j=1}^{\zeta} \phi_X(\zeta, j) X_{\zeta-j} = 0,$$

since $\phi_X(\zeta, j) = 0$ for $j \in \{1, \cdots, \zeta\}$ (see Theorem 2.1). For $t = sk$ in expression (2.12) we have

$$m_{sk} = \sum_{j=1}^{sk} \phi_X(sk, j) X_{sk-j} = \sum_{j=1}^{k} \phi_X(sk, sj) X_{s(k-j)} = \sum_{j=1}^{k} \phi_X(k, j) X_{s(k-j)},$$

since from expression (2.8) one knows that $\phi_X(sk, sj+\zeta) = 0$, for any $\zeta \neq 0$, and $j \in \{1, 2, \cdots, k\}$. For $t = sk + \zeta$ in expression (2.12), where $\zeta \in A$, we have

$$m_{sk+\zeta} = \sum_{j=1}^{sk+\zeta} \phi_X(sk+\zeta, j) X_{sk+\zeta-j} = \sum_{j=1}^{k} \phi_X(sk+\zeta, sj) X_{s(k-j)+\zeta} = \sum_{j=1}^{k} \phi_X(k, j) X_{s(k-j)+\zeta},$$

since from expression (2.8) one knows that $\phi(sk+\xi, sj+\eta) = 0$, for any $\xi, \eta \in A$, and $j \in \{1, 2, \cdots, k\}$. This proves expression (2.10).

Secondly, one wants to prove expression (2.11). For any stationary stochastic process with zero mean, the conditional variance, given the past observations, is given by (see Hosking, 1984)

$$v_t \equiv Var(X_t |\, X_\ell,\, \ell < t) = \sigma_\varepsilon^2 \prod_{j=1}^{t} [1 - \phi_X^2(j, j)], \tag{2.13}$$

where σ_ε^2 is the variance of the white noise process.

From item (v) of Theorem 2.1, $\phi_X(sk+\zeta, sk+\zeta) = 0$, when $\zeta \in A$. Therefore, $v_\zeta = \sigma_\varepsilon^2 \prod_{j=1}^{\zeta}[1 - \phi_X^2(j, j)] = \sigma_\varepsilon^2$. For $t = sk$ in expression (2.13), one has

$$v_{sk} = \sigma_\varepsilon^2 \prod_{j=1}^{sk} [1 - \phi_X^2(j, j)] = \sigma_\varepsilon^2 \prod_{j=1}^{k} [1 - \phi_X^2(sj, sj)].$$

Finally, for $t = sk + \zeta$ in expression (2.13), where $\zeta \in A$, one has

$$v_{sk+\zeta} = \sigma_\varepsilon^2 \prod_{j=1}^{sk+\zeta} [1 - \phi_X^2(j, j)] = \sigma_\varepsilon^2 \prod_{j=1}^{k} [1 - \phi_X^2(sj, sj)] = v_{sk},$$

since $\phi_X(sk+\zeta, sk+\zeta) = 0$, whenever $k \in \mathbb{N}$, and $\zeta \in A$.

Thus, the system (2.11) holds, and we conclude the proof.

□

The next section presents the estimation procedures for the *seasonal differencing parameter D*.

3 ESTIMATION PROCEDURES

In the literature of the stochastic SARFIMA processes, there exist several estimation procedures for the seasonal differencing parameter D. In this section we summarize six different procedures: one in the parametric and five in the semiparametric class.

In the semiparametric class we deal with the regression method using the periodogram function with total number of regressors $g(n) = n^\alpha$, where n is the sample size e $0 < \alpha < 1$. This estimator is proposed by Geweke and Porter-Hudak (1983) and it will be denoted by GPH.

We also consider the method proposed by Reisen (1994), denoted here by SPR. The regression estimator SPR is obtained by replacing the periodogram function by the smoothed periodogram function with Parzen lag window. Reisen (1994) shows that SPR is obtained of the same form as the GPH with truncation point in the Parzen lag window defined by $\nu = n^\beta$, $0 < \beta < 1$.

The regression estimator R, proposed by Robinson (1995), is a modified version of the estimator GPH where the number of regressors $g(n)$ starts from $l > 1$. The trimming value l tends to infinity more slowly than $g(n)$. The regression estimator SR is obtained similarly to the R method, where now the spectral density function is estimated by its smoothed version with Parzen lag window (see Robinson, 1995).

The $GPHTa$ method, also used in the works by Hurvich and Ray (1995), and Velasco (1999) is the fifth semiparametric approach considered here. In this method the modified periodogram function is given by

$$I(w_j) = \frac{1}{\sum_{t=0}^{n-1} h(t)^2} \left| \sum_{t=0}^{n-1} h(t) X_t e^{-iw_j t} \right|^2,$$

where the tapered data is obtained from the cosine-bell function $h(t) = \frac{1}{2}\left[1 - \cos\left(\frac{2\pi(t+0.5)}{n}\right)\right]$. The estimator is then obtained similarly to the GPH method.

The parametric approximated maximum likelihood estimator, proposed by Whittle (1953), involves the function

$$Q(\eta) = \int_{-\pi}^{\pi} \frac{I(w)}{f_X(w;\eta)} dw,$$

where $f_X(\cdot\,;\eta)$ is the spectral density function of a SARFIMA$(0,D,0)_s$ process and η denotes the vector of unknown parameters. The W estimator is the value of η which minimizes the function $Q(\cdot)$. Since σ_ε^2 is setting equal to 1.0, in this case the vector η is given only by the parameter D. More details of this estimator can be found in Whittle (1953) and Fox and Taqqu (1986).

4 FORECASTING ANALYSIS

Let $\{X_t\}_{t\in\mathbb{Z}}$ be a SARFIMA$(0,D,0)_s$ process with $D \in (-0.5, 0.5)$, given by the expression (2.2). Suppose one wants to forecast the value X_{t+h} for h-step ahead. The minimum mean squared error forecasting value is given by

$$\widehat{X}_t(h) \equiv \mathbb{E}\,(X_{t+h} \,|\, X_\ell\,, \ell \leqslant t). \tag{4.1}$$

It minimizes the mean squared error of forecasting $\mathbb{E}(X_{t+h} - \widehat{X}_t(h))$. In this case, the forecasting error is given by

$$e_t(h) \;=\; X_{t+h} \;-\; \widehat{X}_t(h). \tag{4.2}$$

To calculate the forecasting values one uses the following facts

(a) $\mathbb{E}(X_{t+h}|X_\ell\,, \ell \leqslant t) = \begin{cases} X_{t+h}, & \text{if} \quad h \leqslant 0, \\ \widehat{X}_t(h), & \text{if} \quad h > 0, \end{cases}$

(b) $\mathbb{E}(\varepsilon_{t+h}|X_\ell\,, \ell \leqslant t) = \begin{cases} \varepsilon_{t+h}, & \text{if} \quad h \leqslant 0, \\ 0, & \text{if} \quad h > 0. \end{cases}$

Therefore, to calculate the forecasting values one

(a) substitutes the past expectations ($h \leqslant 0$) for known values, X_{t+h} and ε_{t+h};

(b) substitutes the future expectations ($h > 0$) for forecasting values $\widehat{X}_t(h)$ and 0.

The following theorem presents some results for forecasting a future value of a SARFIMA$(0,D,0)_s$ process, given by the expression (2.2).

Theorem 4.1: Let $\{X_t\}_{t\in\mathbb{Z}}$ *be a SARFIMA*$(0,D,0)_s$ *process, with zero mean and seasonality* $s \in \mathbb{N}$*, given in expression (2.2). Consider* $D > -0.5$*. Then, for all* $h \in \mathbb{N}$*,*

(i) *the minimum mean squared error forecasting value is given by*

$$\widehat{X}_n(h) = -\sum_{k\geqslant 1} \pi_k \, \widehat{X}_n(h - sk), \tag{4.3}$$

where π_k *is given in expression (2.3);*

(ii) *the forecasting error is given by* $e_n(h) = \sum_{k=0}^{\lceil \frac{h}{s} \rceil - 1} \psi_k \, \varepsilon_{n+h-sk}$, *where* ψ_k *is given by (2.4) and* $\lceil x \rceil$ *is the smallest integer greater or equal to* x*;*

(iii) *the theoretical and sample variances of the forecast error are given, respectively, by*

$$Var(e_n(h)) = \sigma_\varepsilon^2 \sum_{k=0}^{\lceil \frac{h}{s} \rceil - 1} \psi_k^2, \quad and \quad \widehat{V}ar(e_n(h)) = \widehat{\sigma}_\varepsilon^2 \sum_{k=0}^{\lceil \frac{h}{s} \rceil - 1} \widehat{\psi}_k^2,$$

where $\widehat{\psi}_k$ *is given by (2.4) when* D *is replaced by one of its estimated value, through some of the estimation procedures proposed in Section* 3*;*

(iv) *the bias and the percentage bias to estimate the theoretical variance of the forecasting error are given by*

$$bias(h) = \widehat{V}ar(e_n(h)) - Var(e_n(h))$$

and

$$perbias(h) = \frac{|\widehat{V}ar(e_n(h)) - Var(e_n(h))|}{Var(e_n(h))} \times 100 \ \%;$$

(v) *the mean squared error of forecasting is given by* $msef_n = \frac{1}{h}\sum_{k=1}^{h}(e_n(k))^2$.

(vi) *Moreover, if the process* $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ *is such that* $\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2)$, *for any* $t \in \mathbb{Z}$, *then the* $100\gamma\%$ *confidence interval for* X_{n+h} *is given by*

$$\widehat{X}_n(h) - z_{\frac{\gamma}{2}} \widehat{\sigma}_\varepsilon \left[\sum_{k=0}^{\lceil \frac{h}{s} \rceil - 1} \widehat{\psi}_k^2 \right]^{\frac{1}{2}} \leqslant X_{n+h} \leqslant \widehat{X}_n(h) + z_{\frac{\gamma}{2}} \widehat{\sigma}_\varepsilon \left[\sum_{k=0}^{\lceil \frac{h}{s} \rceil - 1} \widehat{\psi}_k^2 \right]^{\frac{1}{2}},$$

where $z_{\frac{\gamma}{2}}$ *is the value such that* $\mathbb{P}(Z \geqslant z_{\frac{\gamma}{2}}) = \frac{\gamma}{2}$, *with* $Z \sim \mathcal{N}(0, 1)$, *and* $\widehat{\psi}_k$ *is given by the above* (iii) *item.*

Proof:

(i) From Theorem 2.1 items (i) and (ii), a SARFIMA$(0, D, 0)_s$ process can be written as an infinite autoregressive, and infinite moving-average representation. Rewriting these infinite representations for lag $t + h$ one has

$$\varepsilon_{t+h} = \sum_{k\geqslant 0} \pi_k \, X_{t+h-sk}, \tag{4.4}$$

$$X_{t+h} = \sum_{k\geqslant 0} \psi_k \, \varepsilon_{t+h-sk}, \tag{4.5}$$

where π_k and ψ_k are given in the equations (2.3) and (2.4). Since we are interested in the h-step ahead forecasting, from equations (4.4) we have

$$\varepsilon_{t+h} = X_{t+h} + \sum_{k \geqslant 1} \pi_k X_{t+h-sk}.$$

Considering $t = n$ and applying the equality (4.1) in the above equation, and using the conditional expectation properties, one has for all $h \geqslant 1$

$$\begin{aligned} \widehat{X}_n(h) &\equiv \mathbb{E}(X_{n+h} \,|\, X_\ell \,, \ell \leqslant n) = \mathbb{E}(\varepsilon_{n+h} - \sum_{k \geqslant 1} \pi_k \, X_{n+h-sk} | \, X_\ell \,, \ell \leqslant n) \\ &= \mathbb{E}\left(\varepsilon_{n+h} \,|\, X_\ell \,, \ell \leqslant n\right) - \sum_{k \geqslant 1} \pi_k \;\; \mathbb{E}(X_{n+h-sk} | \, X_\ell \,, \ell \leqslant n) \\ &= -\sum_{k \geqslant 1} \pi_k \;\, \widehat{X}_n(h - sk), \end{aligned}$$

since $\mathbb{E}(\varepsilon_{n+h} | \, X_\ell \,, \ell \leqslant n) = \mathbb{E}(\varepsilon_{n+h}) = 0$ and $\widehat{X}_n(j) = X_{n+j}$, for $j \leqslant 0$.

(ii) To obtain the forecasting error $e_n(h)$ one considers the infinite moving-average representation for SARFIMA$(0, D, 0)_s$ processes. Applying the equality (4.1) into the equation (4.5) one obtains, for all $h \geqslant 1$,

$$\begin{aligned} \widehat{X}_n(h) &= \mathbb{E}(X_{n+h} \,|\, X_\ell \,, \ell \leqslant n) = \sum_{k \geqslant 0} \psi_k \, \mathbb{E}(\varepsilon_{n+h-sk} | \, X_\ell \,, \ell \leqslant n) \\ &= \sum_{k \geqslant 0} \psi_k \, \widehat{\varepsilon}_n(h - sk) = \sum_{k \geqslant \lceil \frac{h}{s} \rceil} \psi_k \, \varepsilon_{n+h-sk}, \end{aligned}$$

since $\mathbb{E}(\varepsilon_{n+h-sk} | X_\ell, \ell \leqslant n) = \mathbb{E}(\varepsilon_{n+h-sk}) = 0$ because ε_{n+h-sk} is independent of X_ℓ, for $\ell \leqslant n$, for all $h \geqslant 1$, and $k \in \{0, 1, \cdots, \lceil \frac{h}{s} \rceil - 1\}$, where $\widehat{\varepsilon}_n(j) = 0$, for $j \geqslant 1$, $\widehat{\varepsilon}_n(j) = \varepsilon_{n+j}$, for $j \leqslant 0$, and $\lceil x \rceil$ is the smallest integer greater or equal to x. Therefore, the forecasting error at the origin n for $h \geqslant 1$ steps ahead, is given by

$$e_n(h) = X_{n+h} - \widehat{X}_n(h) = \sum_{k \geqslant 0} \psi_k \varepsilon_{n+h-sk} - \sum_{k = \lceil \frac{h}{s} \rceil}^{\infty} \psi_k \varepsilon_{n+h-sk} = \sum_{k=0}^{\lceil \frac{h}{s} \rceil - 1} \psi_k \, \varepsilon_{n+h-sk}.$$

(iii) The theoretical variance of the forecasting error is given by

$$Var(e_n(h)) = Var\left(\sum_{k=0}^{\lceil \frac{h}{s} \rceil - 1} \psi_k \, \varepsilon_{n+h-sk} \right) = \sum_{k=0}^{\lceil \frac{h}{s} \rceil - 1} \psi_k^2 Var(\varepsilon_{n+h-sk}) = \sigma_\varepsilon^2 \sum_{k=0}^{\lceil \frac{h}{s} \rceil - 1} \psi_k^2.$$

To obtain the theoretical variance of the forecasting error, denoted by $Var(e_n(h))$, one assumes that all parameters in the model are known. In practical situations one uses the estimated model for forecasting. In this case, the sample variance of the forecasting error, denoted by $\widehat{V}ar(e_n(h))$, is obtained by replacing σ_ε^2, and ψ_k, respectively, for their estimators $\widehat{\sigma}_\varepsilon^2$, and $\widehat{\psi}_k$. That is,

$$\widehat{V}ar(e_n(h)) = \widehat{\sigma}_\varepsilon^2 \sum_{k=0}^{\lceil \frac{h}{s} \rceil - 1} \widehat{\psi}_k^2, \quad \text{for all} \quad h \geqslant 1,$$

where $\widehat{\psi}_k$ is given by (2.4) when D is replaced by one of its estimated value.

(iv) The $bias_n(h)$ and $perbias_n(h)$ follow immediately from item (iii). The $perbias_n(h)$ determines the percentage bias due to the estimation of the theoretical variance of the forecasting h-step ahead.

(v) the mean squared error of forecasting at the origin n, denoted by $msef_n$, is obtained through the arithmetic average of the h forecasting squared error. That is,

$$msef_n = \frac{1}{h}\sum_{k=1}^{h}(e_n(k))^2 = \frac{1}{h}\sum_{k=1}^{h}\left(\sum_{\ell=0}^{\lceil \frac{h}{s}\rceil -1} \psi_\ell \, \varepsilon_{n+h-s\ell}\right)^2 .$$

(vi) From the additional hypothesis that $\varepsilon_t \sim \mathcal{N}(0,\sigma_\varepsilon^2)$, for any $t \in \mathbb{Z}$, then the conditional distribution of X_{n+h} given X_ℓ, for $\ell \leqslant n$, is $\mathcal{N}(\widehat{X}_n(h), Var(e_n(h)))$. Hence, the confidence interval for X_{n+h} follows immediately, replacing $Var(e_n(h))$ by $\widehat{V}ar(e_n(h))$.

□

5 SIMULATION RESULTS

In this section we analyze the behavior of the estimators, presented in Section 3. The processes $\{X_t\}_{t\in\mathbb{Z}}$ in equation (2.2) were generated as suggested by Hosking (1984), when $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a Gaussian white noise process with zero mean, and variance $\sigma_\varepsilon^2 = 1.0$. The generating method proposed by Hosking (1984) uses the Durbin-Levinson's algorithm that relates the partial autocorrelation and the autocorrelation functions. In Brietzke et al. (2005) the Durbin-Levinson's algorithm recurrent expression was fully calculated for these processes. This closed formula is based on some properties of the hypergeometric functions and it allows to obtain the partial autocorrelation function of order k for any SARFIMA$(0,D,0)_s$ process. The innovation process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ was generated using the RNNOR subroutine of the IMSL library. The mean μ of the process $\{X_t\}_{t\in\mathbb{Z}}$ was assumed to be zero.

The estimation results were obtained for time series with small and large sample sizes, i.e., for $n = 300$, and $1,000$ respectively, and in both cases the estimators are the average of 500 replications. This number of replications is already good enough because of the ergodicity of the process $\{X_t\}_{t\in\mathbb{Z}}$ (see Remark **(5)** after Theorem 2.1) but we also have results when 1,000 replications were used. These results are available upon request, and they did not show a significant improvement for the parameter estimations.

We have considered several values for D in the range $(-0.5, 0.5)$, i.e., antipersistent processes ($D < 0.0$), and persistent or long memory processes ($D > 0.0$), and also several values for the seasonality period ($s \geqslant 2$). However, we report here only the results for $D = 0.2, 0.4, 0.45$, and $s = 3, 6, 12$ since the pattern is similar for the other cases, and they are available upon request.

For the semiparametric estimators we consider three different methods to choose the total number of regressors $g(n) = n^{0.55}$. They are explained below:

• **Method 1**: The total number of regressors $g(n)$ is divided among all seasonal frequencies for each value of s. For instance, when $s = 2$, the number of regressors $g(n)$ is considered only once in its total value; when $s = 12$ the number of regressors $g(n)$ is divided into six equal parts.

• **Method 2**: The total number of regressors $g(n)$ is considered only at the first seasonal frequency, independently of the s value.

• **Method 3**: The total number of regressors $g(n)$ is considered in the regression range at each seasonal frequency.

The truncation point in the Parzen lag window, for the SPR and SR methods, is $\nu = n^{\beta}$, with $\beta = 0.9$ (see Reisen, 1994 for a discussion on the β value).

All tables include the mean estimated value of the parameter D, depending on the estimation procedure used, and its mean squared error value (mse). The estimator given by a number equal to 1, 2 or 3 in parenthesis means its mean estimated value considering one of the methods, respectively **1**, **2** or **3**, for the choice of regressor's number.

Table 5.1: Results for the SARFIMA$(0, D, 0)_s$ model when $D \in \{0.2, 0.4, 0.45\}$, $n \in \{300, 1000\}$ and $s = 3$.

	$D = 0.2$							
	$n = 300$				$n = 1,000$			
Estimator	**Method 1**	**Method 2**	**Method 3**		**Method 1**	**Method 2**	**Method 3**	
GPH	0.2087	0.1989	0.1989		0.2040	0.1944	0.1944	
mse	0.0639	0.0256	0.0256		0.0300	0.0120	0.0120	
SPR	0.1444	0.1618	0.1618		0.1655	0.1764	0.1764	
mse	0.0432	0.0193	0.0193		0.0168	0.0082	0.0082	
R	0.2108	0.1964	0.1964		0.2099	0.1952	0.1952	
mse	0.1155	0.0360	0.0360		0.0422	0.0149	0.0149	
SR	0.2030	0.1919	0.1919		0.2043	0.1967	0.1967	
mse	0.0573	0.0219	0.0219		0.0200	0.0089	0.0089	
$GPHTa$	0.1900	0.1938	0.1982		0.1626	0.2068	0.1815	
mse	0.0610	0.0610	0.0273		0.0129	0.0129	0.0065	
W				0.1421				0.1816
mse				0.0132				0.0026
	$D = 0.4$							
	$n = 300$				$n = 1,000$			
Estimator	**Method 1**	**Method 2**	**Method 3**		**Method 1**	**Method 2**	**Method 3**	
GPH	0.4122	0.4110	0.4110		0.4134	0.4028	0.4028	
mse	0.0564	0.0213	0.0213		0.0280	0.0123	0.0123	
SPR	0.4231	0.4205	0.4205		0.4070	0.4039	0.4039	
mse	0.0290	0.0135	0.0135		0.0168	0.0085	0.0085	
R	0.4009	0.4060	0.4060		0.4179	0.4027	0.4027	
mse	0.1117	0.0337	0.0337		0.0463	0.0168	0.0168	
SR	0.4524	0.4323	0.4323		0.4343	0.4159	0.4159	
mse	0.0517	0.0199	0.0199		0.0246	0.0105	0.0105	
$GPHTa$	0.4352	0.4156	0.4321		0.4136	0.4032	0.4095	
mse	0.0562	0.0562	0.0264		0.0102	0.0102	0.0062	
W				0.3436				0.3832
mse				0.0106				0.0026
	$D = 0.45$							
	$n = 300$				$n = 1,000$			
Estimator	**Method 1**	**Method 2**	**Method 3**		**Method 1**	**Method 2**	**Method 3**	
GPH	0.4443	0.4444	0.4444		0.4745	0.4636	0.4636	
mse	0.0593	0.0209	0.0209		0.0266	0.0103	0.0103	
SPR	0.5713	0.5330	0.5330		0.5537	0.5157	0.5157	
mse	0.0463	0.0202	0.0202		0.0259	0.0115	0.0115	
R	0.4326	0.4395	0.4395		0.4684	0.4586	0.4586	
mse	0.1101	0.0320	0.0320		0.0404	0.0138	0.0138	
SR	0.5562	0.5148	0.5148		0.5494	0.5059	0.5059	
mse	0.0558	0.0202	0.0202		0.0287	0.0111	0.0111	
$GPHTa$	0.5090	0.4221	0.4758		0.5237	0.4384	0.4964	
mse	0.0513	0.0513	0.0225		0.0152	0.0152	0.0073	
W				0.3753				0.4317
mse				0.0115				0.0021

Tables 5.1 to 5.3 present the estimation results for the SARFIMA$(0, D, 0)_s$ processes when $s \in \{3, 6, 12\}$, and $\alpha = 0.55$. Similar results when $\alpha = 0.65$ are available upon request. One observes that the mean squared error values for the estimates of $D \in (-0.5, 0.5)$ in SARFIMA$(0, D, 0)_s$ processes, with $s \in \{3, 6, 12\}$, and $\alpha = 0.55$, decrease when the sample size increases. For the GPH, SPR, R, SR, and $GPHTa$ estimators, the mean squared error value decreases whatever is the method considered for determining the total number of regressors.

For small seasonal period (for instance, when $s = 3$), the estimator W has the smallest mean squared error value. Nevertheless, the mean squared error value increases whenever the seasonality s increases (for instance, when $s \in \{6, 12\}$), since its bias also increases.

Table 5.3 shows that the W estimator needs larger sample size when the seasonal period increases, in order to achieve the maximum value of the likelihood function $\mathcal{L}_n(\cdot)$.

When $s = 6$, $n = 300$, and $D \in \{0.2, 0.4\}$, $SPR(\mathbf{2})$, and $SPR(\mathbf{3})$ have both the smallest mean squared error value. However, when $D = 0.45$, $GPH(\mathbf{2})$, and $GPH(\mathbf{3})$ are better in the sense of small mean squared error value for this size of n. When $n = 1,000$, and $D \in \{0.2, 0.45\}$ the W estimator is the best one. When $D = 0.4$, $GPHTa(\mathbf{3})$ is outperformed by the others in this sample size case.

Table 5.2: Results for the SARFIMA$(0, D, 0)_s$ model when $D \in \{0.2, 0.4, 0.45\}$, $n \in \{300, 1000\}$ and $s = 6$.

				$D = 0.2$				
	$n = 300$				$n = 1,000$			
Estimator	**Method 1**	**Method 2**	**Method 3**		**Method 1**	**Method 2**	**Method 3**	
GPH	0.1868	0.1953	0.1953		0.2266	0.2070	0.2070	
mse	0.0884	0.0280	0.0280		0.0483	0.0116	0.0116	
SPR	0.1164	0.1644	0.1644		0.1657	0.1851	0.1851	
mse	0.0535	0.0219	0.0219		0.0289	0.0080	0.0080	
R	0.2271	0.2099	0.2099		0.2250	0.2032	0.2032	
mse	0.2129	0.0486	0.0486		0.0804	0.0152	0.0152	
SR	0.2043	0.2038	0.2038		0.2173	0.2041	0.2041	
mse	0.0865	0.0313	0.0313		0.0418	0.0094	0.0094	
$GPHTa$	0.2130	0.2173	0.2225		0.1680	0.1998	0.1826	
mse	0.0819	0.0819	0.0526		0.0121	0.0121	0.0056	
W				0.0733				0.1614
mse				0.0396				0.0062
				$D = 0.4$				
	$n = 300$				$n = 1,000$			
Estimator	**Method 1**	**Method 2**	**Method 3**		**Method 1**	**Method 2**	**Method 3**	
GPH	0.4011	0.3878	0.3878		0.4002	0.4048	0.4048	
mse	0.0920	0.0218	0.0218		0.0357	0.0108	0.0108	
SPR	0.4513	0.4280	0.4280		0.4213	0.4162	0.4162	
mse	0.0369	0.0118	0.0118		0.0194	0.0064	0.0064	
R	0.3848	0.3796	0.3796		0.3946	0.4038	0.4038	
mse	0.2111	0.0356	0.0356		0.0624	0.0126	0.0126	
SR	0.4703	0.4264	0.4264		0.4422	0.4210	0.4210	
mse	0.0835	0.0199	0.0199		0.0341	0.0085	0.0085	
$GPHTa$	0.4143	0.4303	0.4222		0.4316	0.4031	0.4178	
mse	0.0649	0.0649	0.0360		0.0091	0.0091	0.0040	
W				0.2542				0.3599
mse				0.0412				0.0060
				$D = 0.45$				
	$n = 300$				$n = 1,000$			
Estimator	**Method 1**	**Method 2**	**Method 3**		**Method 1**	**Method 2**	**Method 3**	
GPH	0.4281	0.4177	0.4177		0.4860	0.4532	0.4532	
mse	0.1021	0.0220	0.0220		0.0382	0.0094	0.0094	
SPR	0.6891	0.5788	0.5788		0.6310	0.5329	0.5329	
mse	0.0856	0.0263	0.0263		0.0497	0.0120	0.0120	
R	0.4003	0.4064	0.4064		0.4728	0.4431	0.4431	
mse	0.2476	0.0390	0.0390		0.0645	0.0127	0.0127	
SR	0.6536	0.5343	0.5343		0.6172	0.5102	0.5102	
mse	0.1045	0.0224	0.0224		0.0542	0.0099	0.0099	
$GPHTa$	0.5012	0.5091	0.5063		0.5539	0.4332	0.4992	
mse	0.0590	0.0590	0.0327		0.0172	0.0172	0.0057	
W				0.2865				0.4067
mse				0.0448				0.0052

For the case when $s = 12$, the best estimator in the smallest mean squared error sense is $GPHTa(\mathbf{3})$ except when $n = 1,000$, and $D = 0.45$, where $GPH(\mathbf{2})$, and $GPH(\mathbf{3})$ outperform the tapering data estimator.

After we estimate the seasonal fractionally differencing parameter D, in order to obtain the h-step ahead forecasting one needs to proceed as follows. From Theorem 4.1, item (i), one needs to truncate the expression for the minimum mean squared error forecasting value, that is,

$$\widehat{X}_n(h) = -\sum_{j=1}^{k} \widehat{\pi}_j \; \widehat{X}_n(h - sj) = -\sum_{j=1}^{k} \widehat{\pi}_j \; \widehat{X}_{n+h-sj}, \tag{5.1}$$

where $k = \left\lfloor \frac{n+h-1}{s} \right\rfloor$, and $\widehat{\pi}_j$ is given by (2.3) when D is replaced by one of its estimated value, through some of the estimation procedures proposed in Section 3.

The forecasting subroutine also includes the forecasting error, the theoretical, and sample variances for the error, the bias and percentage bias obtained at step h when one estimates the theoretical variance h-step ahead, and the mean squared error of forecasting at step h (see Theorem 4.1).

Table 5.3: Results for the SARFIMA$(0, D, 0)_s$ model when $D \in \{0.2, 0.4, 0.45\}$, $n \in \{300, 1000\}$ and $s = 12$.

$D = 0.2$								
	$n = 300$				$n = 1,000$			
Estimator	**Method 1**	**Method 2**	**Method 3**		**Method 1**	**Method 2**	**Method 3**	
GPH	0.1914	0.1831	0.1831		0.2111	0.2109	0.2109	
mse	0.2563	0.0636	0.0636		0.0989	0.0143	0.0143	
SPR	0.0169	0.1427	0.1427		0.1009	0.1802	0.1802	
mse	0.1075	0.0512	0.0512		0.0625	0.0102	0.0102	
R	0.2147	0.1843	0.1843		0.1844	0.2077	0.2077	
mse	1.4583	0.1375	0.1375		0.2493	0.0194	0.0194	
SR	0.1281	0.1932	0.1932		0.1576	0.2018	0.2018	
mse	0.2005	0.0809	0.0809		0.1006	0.0125	0.0125	
$GPHTa$	0.1796	0.1878	0.1433		0.1428	0.2009	0.1643	
mse	0.2270	0.2270	0.0246		0.0217	0.0217	0.0084	
W				-0.0740				0.1305
mse				0.1248				0.0147

$D = 0.4$								
	$n = 300$				$n = 1,000$			
Estimator	**Method 1**	**Method 2**	**Method 3**		**Method 1**	**Method 2**	**Method 3**	
GPH	0.3787	0.3854	0.3854		0.4166	0.4049	0.4049	
mse	0.2543	0.0565	0.0565		0.0972	0.0125	0.0125	
SPR	0.4751	0.4710	0.4710		0.4600	0.4264	0.4264	
mse	0.0417	0.0314	0.0314		0.0328	0.0077	0.0077	
R	0.3212	0.3822	0.3822		0.4223	0.4032	0.4032	
mse	1.2589	0.1270	0.1270		0.2112	0.0175	0.0175	
SR	0.5221	0.4761	0.4761		0.5058	0.4257	0.4257	
mse	0.1455	0.0590	0.0590		0.0795	0.0113	0.0113	
$GPHTa$	0.3982	0.4358	0.4343		0.4671	0.4295	0.4281	
mse	0.1521	0.1521	0.0138		0.0156	0.0156	0.0055	
W				0.1204				0.3247
mse				0.1213				0.0152

$D = 0.45$								
	$n = 300$				$n = 1,000$			
Estimator	**Method 1**	**Method 2**	**Method 3**		**Method 1**	**Method 2**	**Method 3**	
GPH	0.4615	0.4214	0.4214		0.4542	0.4415	0.4415	
mse	0.2611	0.0591	0.0591		0.0925	0.0107	0.0107	
SPR	0.8217	0.6796	0.6796		0.7083	0.5489	0.5489	
mse	0.1632	0.0699	0.0699		0.0871	0.0147	0.0147	
R	0.4810	0.4139	0.4139		0.4687	0.4413	0.4413	
mse	1.4100	0.1253	0.1253		0.2333	0.0160	0.0160	
SR	0.9084	0.6466	0.6466		0.7155	0.5200	0.5200	
mse	0.3012	0.0758	0.0758		0.1206	0.0126	0.0126	
$GPHTa$	0.5006	0.5500	0.5588		0.6006	0.5192	0.5287	
mse	0.1383	0.1383	0.0202		0.0302	0.0302	0.0090	
W				0.1513				0.3598
mse				0.1358				0.0152

These calculations were done for the total number of replicated time series of size n, and after that the arithmetic average was calculated for each one of the h-steps. In this work we are interested in 5-step ahead forecasting. Figures 5.1 to 5.3 present the $msef_n$ when the seasonality $s \in \{3, 6, 12\}$, respectively, for $n = 300$ and $1,000$ and $D = 0.2$. Here we consider only **Methods 1** and **3** since the results when using **Method 2** are similar to **Method 3**. All three methods were considered only for the estimator $GPHTa$. We again report here the results when $\alpha = 0.55$ for all semiparametric estimators.

Figures 5.1 to 5.3 show the mean squared error value of forecasting at the origin n for two different sample sizes n ($n \in \{300; 1,000\}$) when $D = 0.2$, and $s \in \{3, 6, 12\}$. When $s = 12$, since the need of more number of regressors is evident, we consider only the case where $n = 1,000$ (see Figure 5.3).

From these figures one observes that the sample forecasting error is close to zero, no matter what horizon for the forecasting step-ahead one is using and for any estimation procedure considered in Section 3. Since the theoretical variance for the forecasting error was set equal to 1.0, we observe that the $msef_n$ should also be approximately equal to 1.0. We remark that the value of the mean squared error value of the forecasting, for any estimation procedure, decreases when the sample size increases.

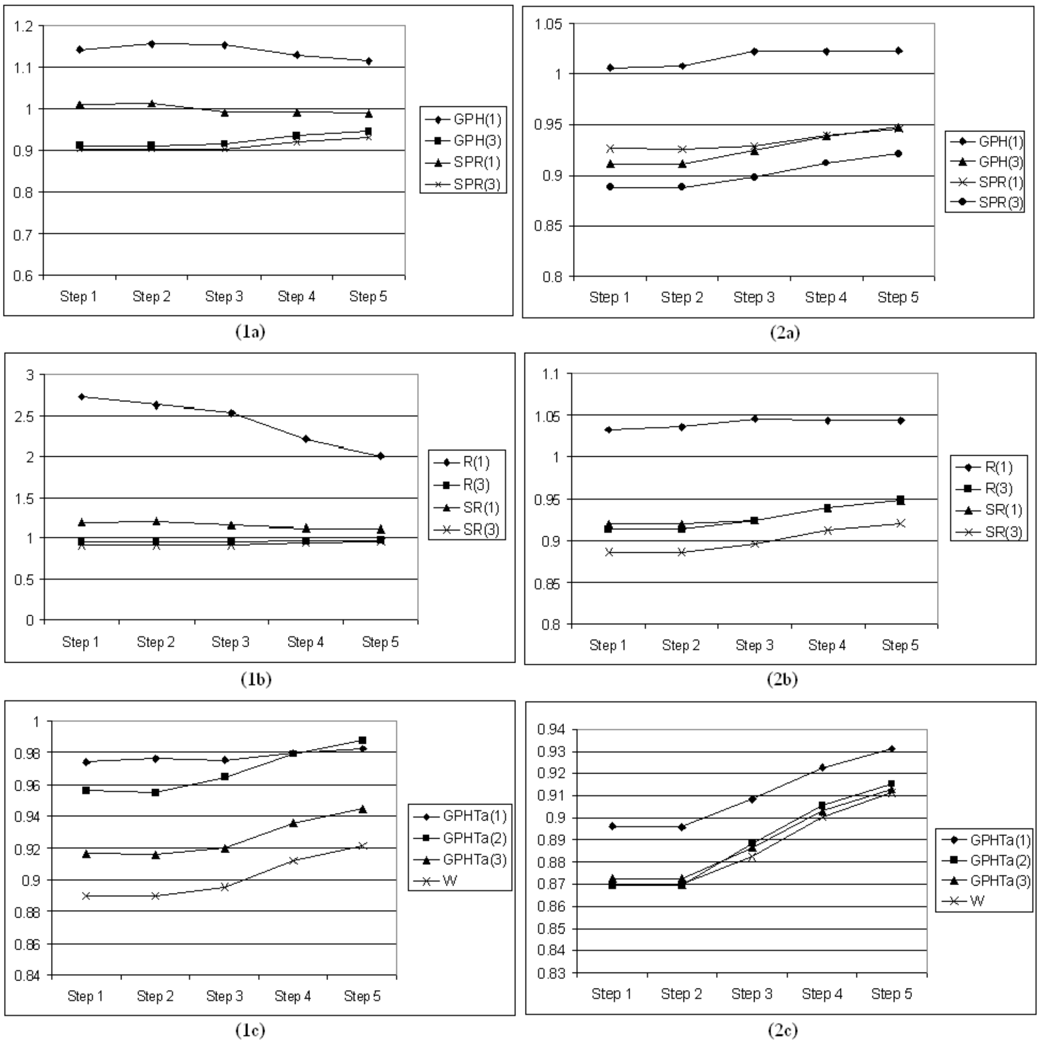


Figure 5.1: The graphics (1a), (1b), and (1c) present the $msef_n$ of time series with $s = 3$, $n = 300$, and $D = 0.2$. Graphics (2a), (2b), and (2c) present the $msef_n$ of time series with $s = 3$, $n = 1,000$, and $D = 0.2$.

Figures 5.1 to 5.3 also show that the $msef_n$ for all estimators considered are very small except for the estimator $R(\mathbf{1})$ where its $msef_n$ is superior to 1.0, and it increases when the value of the seasonality increases. Again for the estimator $R(\mathbf{1})$, the value of the mean squared error of forecasting is the largest one, independently of the values of s, n, and α.

We also remark here that the forecasting results have a great improvement when the value of α increases, i.e., for large values of α, the forecasting error, and the mean squared error of forecasting decrease. We do not report these results here but they are available upon request.

We also consider different values of replications. In fact, when one considers $1,000$ replications instead of 500, the results for the $msef_n$ do not significantly change.

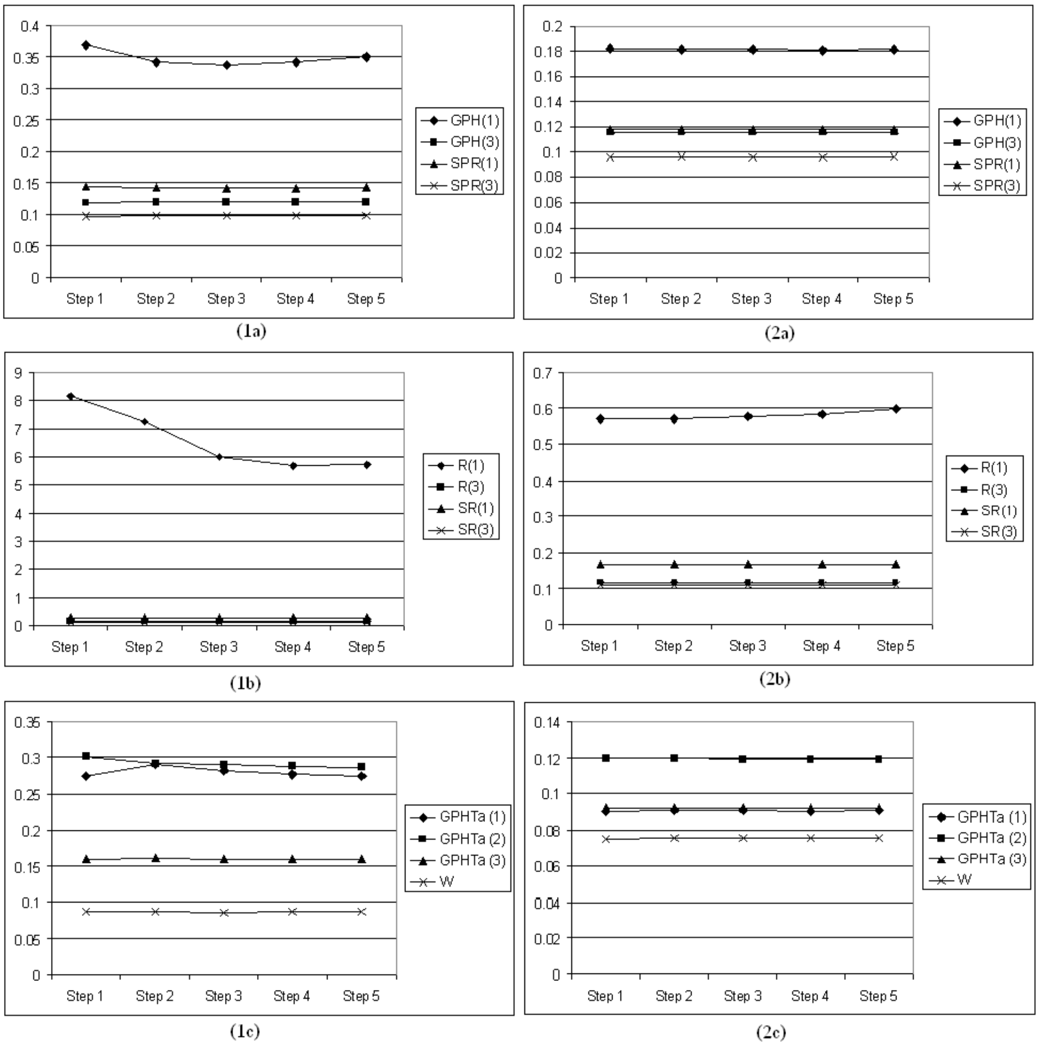


Figure 5.2: The graphics (1a), (1b), and (1c) present the $msef_n$ of time series with $s = 6$, $n = 300$, and $D = 0.2$. Graphics (2a), (2b), and (2c) present the $msef_n$ of time series with $s = 6$, $n = 1,000$, and $D = 0.2$.

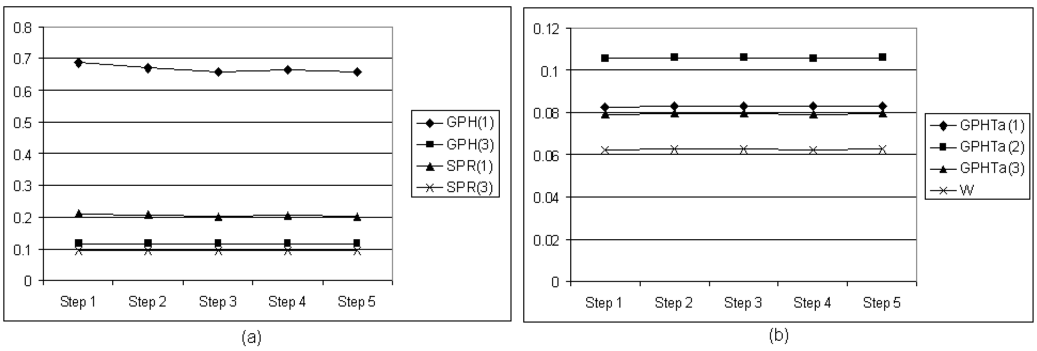


Figure 5.3: The graphics (a) and (b) present the $msef_n$ of time series with $s = 12$, $n = 1,000$, and $D = 0.2$.

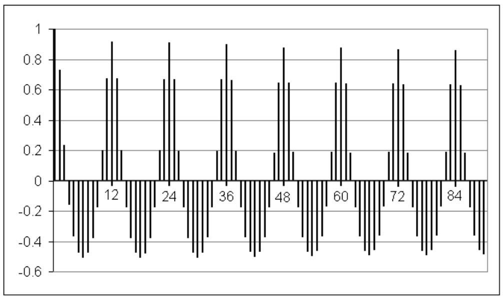
6 APPLICATIONS

In this section we analyze an observed time series data, and also a simulated seasonal fractionally integrated ARMA time series. Our goal is to analyze these two time series in order to detect whether seasonal long memory is present in the data.

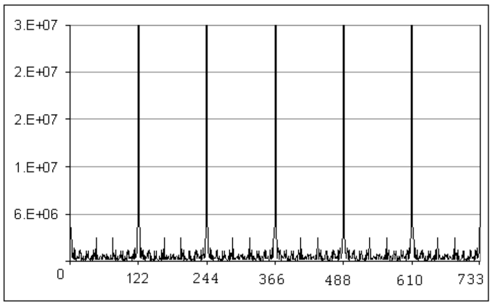
6.1 Nile River Monthly Flows Data

We consider the time series Nile River monthly flows at Aswan kindly provided by A. Montanari (for the graphic of the data see Montanari et al., 2000). It consists of 1,466 observations, from August of 1872 to September of 1994, and it is approximately a Gaussian time series. Figures 6.1 (a) and (b) present, respectively its sample autocorrelation, and periodogram functions.

Figures 6.1 (a), and (b) show long memory features for this time series, since its sample autocorrelation has a slowly hyperbolic decay, and its periodogram function exhibits periodic pattern caused by an annual cycle. Figure 6.1 (b) shows the peaks on the Fourier frequencies w_j, where $j = [n/s]\, i = [1,466/12]\, i = 122i$, for $i = 0, 1, \cdots, 6$. These features are also reported in Montanari et al. (2000).



(a)



(b)

Figure 6.1: The graphs are related to the time series Nile River Monthly Flows data at Aswan: (a) sample autocorrelation function and (b) periodogram function.

The best model fitted for the original data was a SARFIMA$(p, d, q) \times (P, D, Q)_s$ with $p = q = P = Q = 1$, $d = 0$, $D = W$ and $s = 12$, where the long memory parameter was estimated by the maximum likelihood method proposed by Fox and Taqqu (1986) (see Section 3), with $W = 0.1980$.

Remark: Even though the best model fitted for the data was a SARFIMA$(p, d, q) \times (P, D, Q)_s$ process, we analised the Nile River monthly flows since it is difficult to find a time series that is modelled by a SARFIMA$(0, D, 0)_s$ process.

Table 6.1: Estimated Values of D for: (a) Nile River Monthly Flows Data and (b) Simulated Time Series Data.

SARFIMA$(0, D, 0)_s$ with $s = 12$ and $\alpha = 0.55$								
	(a) Nile River Monthly Flows Data				(b) Simulated Time Series Data			
Estimator	**Method 1**	**Method 2**	**Method 3**		**Method 1**	**Method 2**	**Method 3**	
GPH	0.2549	0.2399	0.2491		0.4443	0.4219	0.4251	
SPR	0.6281	0.3126	0.3234		0.4149	0.4216	0.3911	
R	0.2154	0.2381	0.2472		0.4511	0.4398	0.4261	
SR	0.5385	0.3046	0.3140		0.4385	0.4071	0.4188	
$GPHTa$	0.5625	0.4196	0.5714		0.4431	0.4185	0.4306	
FT				0.1980				0.3834

Table 6.1 (a) gives the estimation results for this time series with seasonality $s = 12$, since Figures 6.1 (a) and (b) exhibits this periodic pattern. We consider all three methods to perform the estimation procedures.

Table 6.2 gives the estimators and its standard deviation (denoted by Std. Dev.) values for the parameters in the SARFIMA$(p, d, q) \times (P, D, Q)_s$ model, that best fitted the Nile River monthly flows data at Aswan.

The residual analysis was also performed for the fitted model and it indicates that the errors are approximately Gaussian white noise.

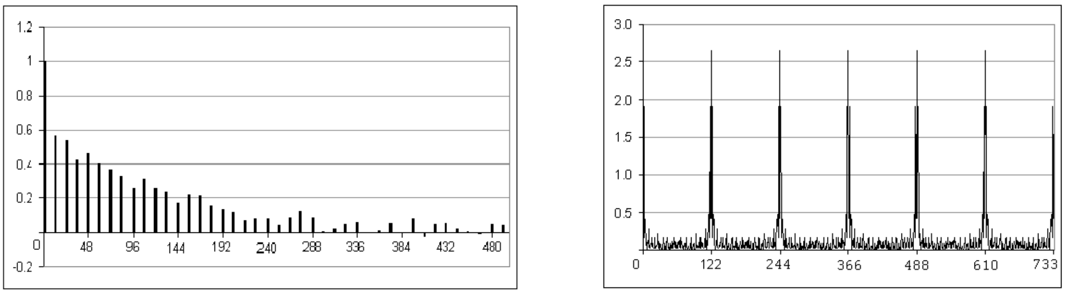
Table 6.2: Fitted Model for the Nile River Flows Data.

SARFIMA$(p,d,q)\times(P,D,Q)_s$ with $p=q=P=Q=1$, $D=W$ and $s=12$					
	ϕ_1	Φ_1	D	θ_1	Θ_1
Estimator	0.6147	0.9944	0.1980	-0.2238	0.9207
Std. Dev.	0.0291	0.0295	0.0011	0.0357	0.0145

6.2 Simulated Time Series Data

Since it is difficult to find a time series that is modelled by a pure SARFIMA$(0,D,0)_s$ process, we consider here a complete estimation, and the forecasting analysis for a simulated seasonal fractionally integrated time series as in expression (2.2), when $n=1,466$, $D=0.4$, and $s=12$.

Figures 6.2 (a), and (b) show the sample autocorrelation, and the periodogram functions of this simulated time series. One observes, from these figures, that there exist long memory characteristics in this time series. Analyzing the periodogram function we also observe a periodic pattern with seasonality $s=12$.



(a) (b)

Figure 6.2: The graphs are related to the simulated time series data: (a) sample autocorrelation function and (b) periodogram function.

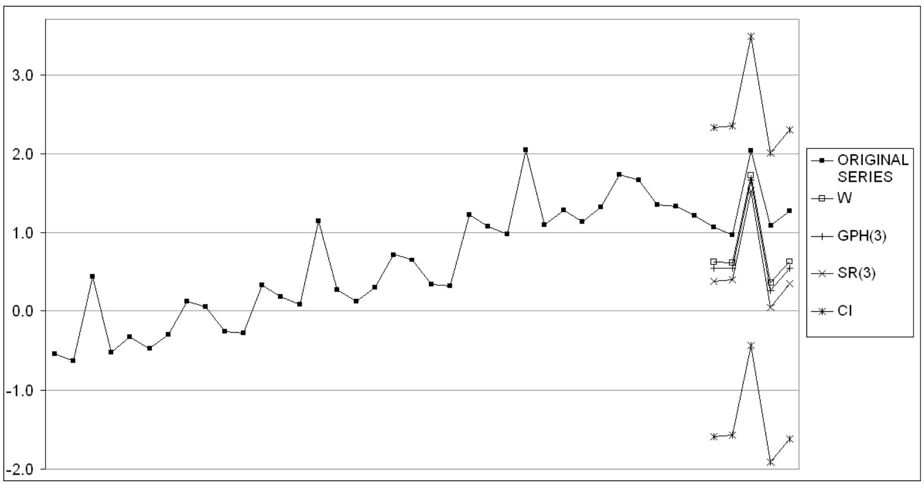


Figure 6.3: Confidence interval at 95% confidence level for the 5-step ahead forecasting in the simulated time series data.

Table 6.1 (b) gives the estimators for the parameter D considering all three methods, for a SARFIMA$(0,D,0)_s$ with $s=12$, that best fitted the simulated time series.

The best estimator for the simulated time series is $W = 0.3834 \simeq 0.4$. In the semiparametric class, **Method 1** always overestimates the parameter D while **Method 3** gives the smallest estimated value for the first three procedures. **Method 2** gives the smallest estimated value only for SR and $GPHTa$. Observe that $SPR(\mathbf{3}) = 0.3911$ is also a very good estimator for D.

Figure 6.3 gives the confidence interval at 95% confidence level for the 5-step ahead forecasting values based on all estimation procedures considered in Section 3 for the simulated time series data.

7 CONCLUSIONS

In studying SARFIMA$(0, D, 0)_s$ processes we emphasize Theorems 2.2 and 4.1. Theorem 2.2 presents the conditional expectation, and the conditional variance for that process. This theorem is very important for generating any SARFIMA$(0, D, 0)_s$ process. Theorem 4.1 gives some properties for forecasting the value X_{n+h}, when $h \geqslant 1$, in SARFIMA$(0, D, 0)_s$ processes.

The performance of five semiparametric and one parametric procedures for estimating the seasonal fractionally differencing parameter D were investigated. In a SARFIMA$(0, D, 0)_s$ model, when $D \in (-0.5, 0.5)$, and $s \geqslant 2$ we observe that the estimation of D can be affected when there exists seasonality present in the model. We also observe that, for almost all cases considered here, the estimators overestimate the theoretical variance.

We show here (see also Bisognin, 2003) that the maximum value of the likelihood function $\mathcal{L}_n(\cdot)$ is only achieved in large enough sample sizes whenever the seasonal period increases. This is a drawback for the W estimator when large seasonal periods are present in the data.

We observe that when the number of replications is doubled from 500 to $1,000$, the results do not significantly improve. We also observe that when the value of α increases (say, from 0.55 to 0.65), the mean squared error values decrease in almost all cases considered here. This was also reported in Lopes et al. (2004) for the case when $s = 1$, where the authors suggest $\alpha \in \{0.6, 0.7, 0.8\}$. The paper by Porter-Hudak (1990) suggests $\alpha \in \{0.62, 0.75\}$ for the case when $s = 12$, and $n = 352$.

Among the different methods considered here for determining the total number of regressors in the semiparametric class, **Method 1**, in general, overestimates the true value of D. **Methods 2**, and **3** have better behaviour for the GPH, SPR, R, and SR estimators, in the sense of small mean squared error value. While **Method 3** determines the best total number of regressors for the $GPHTa$ estimator.

We also conclude that when the parameter D is getting close to the non-stationary region (for instance, when $D = 0.45$), the estimators SPR, SR, $GPHTa$, and W have larger bias compared to the GPH, and R estimators.

Observing the mean value of each estimator we conclude that the GPH procedure has the tendency of overestimating the true parameter value, except for the case when $s = 12$, $n = 300$, and $D \in \{0.2, 0.4\}$. The SPR procedure has the tendency of underestimating the true parameter value except when $s \in \{6, 12\}$, and $D = 0.4$. Only **Method 3** underestimates the true parameter value when $n = 1,000$. The SPR estimator has always the best performance in the sense of smaller mean squared error value. This was already expected in comparison with ARFIMA models (when $s = 1$). We also observe that the SPR estimation procedure is outperformed by the GPH only when $s = 3$ for all three methods considered here and also always when $D = 0.2$. Generally, **Method 2** provides smaller mean squared error values for the estimators while **Method 3** provides the smallest one when one compares all three methods considered here. All estimators improve as the sample size increases, as was expected.

When we performed a 5-step ahead forecasting analysis for this process, we observed that the sample forecasting error is very close to zero and the estimators overestimate the theoretical variance, independently of the estimation procedure considered here. The estimator $R(\mathbf{1})$ is the one with larger mean squared error of forecasting, and it increases with the seasonality. However, the estimator W has the smallest mean squared error of forecasting, and it remains almost the same no matter how large is the seasonal value.

If one relates the forecasting results with the α value, one observes that when it increases, the results improve significantly, that is, for large values of α, the forecasting error decreases, the sample variance of the forecasting error approaches to the theoretical variance, and, consequently, the bias, and the percentage bias decrease. The mean squared error of forecasting also decreases. This is also true when the sample size increases.

We applied the methodology to the Nile River monthly flows at Aswan, and observed that a SARFIMA$(p,d,q)\times(P,D,Q)_s$ model, with $p=q=P=Q=1$, $d=0$, $D=0.1980$ and $s=12$, fitted well the original data.

Since we did not find an observed time series modelled by a SARFIMA$(0,D,0)_s$ process, we consider a simulated seasonal fractionally integrated one to give a complete estimation and forecasting analysis.

For future work we want to consider the performance of these estimation procedures, and also the forecasting analysis when small and large short-run parameters are included in the process.

ACKNOWLEDGEMENTS

Many thanks are due to Alberto Montanari for having provided the series of the monthly flows data of the Nile River. C. Bisognin was supported by CAPES-Brazil. S.R.C. Lopes was partially supported by CNPq-Brazil, by Pronex *Probabilidade e Processos Estocásticos* (Convênio MCT/CNPq/FAPERJ-Edital 2003), by CNPq 019/476781/2004-3 *Modelos com Dependência de Longo Alcance: Análise Probabilística e Inferência* and also by *Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul* (FAPERGS).

REFERENCES

Beran J. 1994. *Statistics for Long-Memory Process.* Chapman and Hall: New York.

Bisognin C. 2003. Estimação e Previsão em Processos com Longa Dependência Sazonal. Master Dissertation from Mathematics Master Program at UFRGS, Porto Alegre, RS, Brazil. URL's address: *http://www.mat.ufrgs.br/∼slopes*.

Bisognin C, Lopes SRC. 2005. Estimating and forecasting the long memory parameter in the presence of periodicity. Technical Report Series A, nº 60, Instituto de Matemática, UFRGS. URL's address: *http://www.mat.ufrgs.br/∼slopes*.

Brietzke EHM, Lopes SRC, Bisognin C. 2005. A closed formula for the Durbin-Levinson´s algorithm in seasonal fractionally integrated processes. To appear in *Mathematical and Computer Modelling.*

Fox R, Taqqu MS. 1986. Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *The Annals of Statistics* **14**: 517-532.

Geweke J, Porter-Hudak S. 1983. The estimation and application of long memory time series model. *Journal of Time Series Analysis* **4**: 221-238.

Granger CWJ, Joyeux R. 1980. An introduction to long memory time series models and fractional differencing. *Journal of Time Series Analysis* **1**(1): 15-29.

Hassler U. 1994. (Mis)specification of long memory in seasonal time series. *Journal of Time Series Analysis* **15**(1): 019-030.

Hosking J. 1981. Fractional differencing. *Biometrika* **68**: 165-167.

Hosking J. 1984. Modelling persistence in hydrological time series using fractional differencing. *Water Resources Research* **20**: 1898-1908.

Hurvich CM, Ray BK. 1995. Estimation of the memory parameter for nonstationary or noninvertible fractionally integrated processes. *Journal of Time Series Analysis* **16**: 017-042.

Lopes SRC, Olbermann BP, Reisen VA. 2004. A comparison of estimation methods in non-stationary ARFIMA processes. *Journal of Statistical Computation and Simulation* **74**(5): 339-347.

Montanari A, Rosso R, Taqqu MS. 2000. A seasonal fractional ARIMA Model applied to the Nile River monthly flows at Aswan. *Water Resources Research* **36**(5): 1249-1259.

Olbermann BP, Lopes SRC, Reisen VA. (2005) Invariance of the first Difference in ARFIMA models. Accepted by publication in *Computational Statistics*.

Ooms M. 1995. Flexible seasonal long memory and economic time series. Preprint of the Econometric Institute, Erasmus University, Rotterdam.

Peiris MS, Singh N. 1996. Predictors for seasonal and nonseasonal fractionally integrated ARIMA models. *Biometrika* **38**(6): 741-752.

Porter-Hudak S. 1990. An application of the seasonal fractionally differenced model to the monetary aggregates. *Journal of American Statistical Association* **85**: 338-344.

Ray BK. 1993. Long-Range forecasting of IBM product revenues using a seasonal fractionally differenced ARMA model. *International Journal of Forecasting* **9**: 255-269.

Reisen VA, Lopes SRC. 1999. It disappears simulations and applications of forecasting long-memory time series models. *Journal of Statistical Planning and Inference* **80**(2): 269-287.

Reisen VA. 1994. Estimation of the fractional difference parameter in the ARIMA (p,d,q) model using the smoothed periodogram. *Journal of Time Series Analysis* **15**(3): 335-350.

Robinson PM. 1995. Log-periodogram regression of time series with long range dependence. *Annals of Statistics* **23**: 1048-1072.

Sowell F. 1992. Maximum Likelihood Estimation of Stationary Univariate Fractionally Integrated Time Series Models. *Journal of Econometrics* **53**: 165-188.

Velasco C. 1999. Non-stationary log-periodogram regression. *Journal of Econometrics* **91**: 325-371.

Whittle P. 1953. Estimation and information in stationary time series. *Arkiv fur Mathematik* **2**: 423-434.

Authors' biographies:

Sílvia R. C. Lopes is a Full Professor in the Statistics Department at UFRGS, Porto Alegre, RS, Brazil. Her research interests include time series analysis, forecasting, long-memory and econometric models, stochastic processes and ergodic theory.

Cleber Bisognin is a Ph.D. Student in the Mathematical Graduate Program at UFRGS, Porto Alegre, RS, Brazil. His research interests include time series analysis, forecasting, long-memory and stochastic processes.

Authors' addresses:

Sílvia R. C. Lopes, Statistics Department and Mathematical Graduate Program, UFRGS, Porto Alegre, RS, Brazil.

Cleber Bisognin, Mathematical Graduate Program, UFRGS, Porto Alegre, RS, Brazil.