Estimation of Parameters in ARFIMA Processes: A Simulation Study

Valderio Reisen
Departamento de Estatistica
UFES
Vitoria - ES
Brazil

Bovas Abraham
Department of Statistics and Actuarial Science
University of Waterloo
Waterloo, Ontario N2L 3G1
Canada

Silvia Lopes
Instituto de Matematica
UFRGS
Porto Alegre - RS
Brazil

Keywords: Fractional differencing, long memory, smoothed periodogram regression, periodogram regression, Whittle maximum likelihood procedure.

Abstract

It is known that, in the presence of short memory components, the estimation of the fractional parameter d in an Autoregressive Fractionally Integrated Moving Average, ARFIMA(p, d, q), process has some difficulties (see (1)). In this paper, we continue the efforts made by Smith et al. (1) and Beveridge and Oickle (2) by conducting a simulation study to evaluate the convergence properties of the iterative estimation procedure suggested by Hosking (3). In this context we consider some semiparametric approaches and a parametric method proposed by Fox-Taqqu (4). We also investigate the method proposed by Robinson (5) and a modification using the smoothed periodogram function.

Acknowledgements: The author S. Lopes thanks the partial support from FAPERGS, Porto Alegre, RS.

1. Introduction

The autoregressive fractionally integrated moving average, ARFIMA(p, d, q), process has widely been used in different fields such as astronomy, hydrology, mathematics and computer science, to represent a time series with long memory property (6). Recently a wide range of estimators for the fractional parameter d have appeared in the time series literature (see for instance, (7), (8), (9), (10), (5,11), (12), (13), (14), (15), (16), (17), (18) and (19). In general, the estimators of d can be categorized into two groups - parametric and semiparametric methods. Within the first group the methods proposed by (4) and (20), which involve the likelihood function, are the most common. In the latter, the most popular, usually referred to as the GPH method, was proposed by Geweke and Porter-Hudak (see (21)); more recently, a modified form of this, was given in (5).

All the parameters (autoregressive, moving average and differencing) can be simultaneously estimated in the parametric approach. In the semiparametric methods, the parameters are estimated in two steps: only d is estimated in the first step and the others are estimated in the second step.

Since Gaussian parametric estimates for long memory range dependent time series models have rigorously been justified by (4), (22), (20), (23) and others, they provide an attractive alternative to the semiparametric methods. However, the Gaussian parametric methods require a great deal of computation while the semiparametric procedures are easy to implement.

The main goal of this paper is to compare the performance of estimating all the parameters of an ARFIMA process based on the algorithm in (3) with that of the parametric Whittle estimator (see(4)). For this analysis we consider several estimators of d which are summarized in Section 2. Section 3 describes the algorithm to estimate the parameters. Section 4 presents the results of a simulation study and Section 5 gives a summary and some concluding remarks.

2. The ARFIMA(p,d,q) model

We now summarize some results for the ARFIMA(p, d, q) model with emphasis on the estimation of the differencing parameter d. Consider the simple ARFIMA(p, d, q) model of the form

$$\Phi(B)(1-B)^d X_t = \Theta(B)\epsilon_t, \quad \text{for} \quad d \in (-0.5, 0.5), \tag{2.1}$$

where $\{\epsilon_t\}$ is a white noise process with $E(\epsilon_t) = 0$ and variance σ_ϵ^2 and B is the back-shift operator such that $BX_t = X_{t-1}$.
The polynomials $\Phi(B) = 1-\phi_1 B - \cdots - \phi_p B^p$ and $\Theta(B) = 1-\theta_1 B - \cdots - \theta_q B^q$ have orders p and q respectively with all their roots outside the unit circle. In this paper we assume that $\{X_t\}$ is a linear process without a deterministic term. We now define $U_t = (1-B)^d X_t$, so that $\{U_t\}$ is an ARMA(p, q) process. The process defined in (2.1) is stationary and invertible (see (3)) and its spectral density function, $f_X(w)$, is given by

$$f_X(w) = f_U(w)(2\sin(w/2))^{-2d}, \quad w \in [-\pi, \pi], \tag{2.2}$$

where $f_U(w)$ is the spectral density function of the process $\{U_t\}$.

2.1. Estimation of d

Now we consider five alternative estimators of the parameter d. Four of them are semiparametric and are based on regression equations constructed from the logarithm of the expression in (2.2). The other one is a parametric method proposed by (4). The methods are summarized as follows:

Periodogram Estimator ($\hat{d}_p$)

The first one denoted by $\hat{d}_p$, was proposed by Geweke and Porter-Hudak (21) who used the periodogram function $I(w)$ as an estimate of the spectral density function in expression (2.2). The number of observations in the regression equation is a function $g(n)$ of the sample size n where $g(n) = n^\alpha, 0 < \alpha < 1$.

Smoothed Periodogram Estimator ($\hat{d}_{sp}$)

The second estimator, denoted by $\hat{d}_{sp}$ in the sequel, was suggested by Reisen (9). This regression estimator is obtained by replacing the spectral density function in the expression (2.2) by the smoothed periodogram function with the Parzen lag window. In this method, $g(n)$ is chosen as above and the truncation point in the Parzen lag window is $m = n^\beta, 0 < \beta < 1$. The appropriate choice of α and β were investigated by (21) and (9), respectively.

Robinson Estimator ($\hat{d}_{pr}$)

The third one is the GPH estimator with mild modifications suggested by Robinson (5), denoted hereafter by $\hat{d}_{pr}$. This estimator regresses $\{\ln I(w_i)\}$

on $\ln(2\sin(w_i/2))^2$, for $i = l, l+1, \cdots, g(n)$, where l is the lower truncation point which tends to infinity more slowly than $g(n)$. Robinson derived some asymptotic results for $\hat{d}_{pr}$, when $d \in (-0.5, 0.5)$, and showed that this estimator is asymptotically less efficient than a Gaussian maximum likelihood estimator of d. Our choice of the bandwidth $g(n)$ is now based on the formulae derived in (24) (page 445), which is optimal in the sense that asymptotically it minimizes the mean squared error of the unlogged periodogram estimator (see (14)). The function $g(n)$ is given by

$$g(n) = \begin{cases} A(d,\tau)n^{\frac{2\tau}{2\tau+1}}, & 0 \leq d \leq 0.25 \\ A(d,\tau)n^{\frac{\tau}{\tau+1-2d}}, & 0.25 < d \leq 0.5 \end{cases}$$

where τ and $A(d,t)$ need to be chosen appropiately. This bandwidth cannot be computed in practice, since it requires knowledge of the true parameter d. However, this problem can be turned around by replacing the unknown parameter d in the $g(n)$ function by either the estimate $\hat{d}_p$ or $\hat{d}_{sp}$. We use this $g(n)$ since it satisfies the conditions $\frac{g(n)}{n} \to 0$ and $g(n)\frac{\ln g(n)}{n} \to 0$ as $g(n)$ and n go to infinity (see condition 1 in (16)). The appropriate choice of the optimal $g(n)$ has been the subject of many papers such as (16) and (17).

Robinson's estimator based on the smoothed periodogram ($\hat{d}_{spr}$)

We suggest, without any mathematical proof, the use of the smoothed periodogram function, with the Parzen lag window, to replace the periodogram in the Robinson's estimator. The truncation point is the same as the one chosen for $\hat{d}_{sp}$ and the number of observations in the regression equation is also the same as the one chosen for $\hat{d}_{pr}$.

Whittle estimator ($\hat{d}_W$)

The fifth estimator is a parametric procedure due to Whittle (see (25)) with modifications suggested by (4) and will be denoted hereafter by $\hat{d}_W$. The estimator $\hat{d}_W$ is based on the periodogram and it involves the function

$$Q(\zeta) = \int_{-\pi}^{\pi} \frac{I(w)}{f_X(w,\zeta)} dw, \tag{2.3}$$

where $f_X(w,\zeta)$ is the known spectral density function at frequency w and ζ denotes the vector of unknown parameters. The Whittle estimator is the value of ζ which minimizes the function $Q(\cdot)$. For the ARFIMA (p,d,q) process the vector ζ contains the parameter d and also all the unknown autoregressive and moving average parameters. For more details see (4), (23) and (6). For computational purposes the estimator $\hat{d}_W$ is obtained by using the discret form of $Q(\cdot)$, as in Dahlhaus (23) (page 1753), that is,

$$\mathcal{L}_n(\zeta) = \frac{1}{2n} \sum_{j=1}^{n-1} \left\{ \ln f_X(w_j,\zeta) + \frac{I(w_j)}{f_X(w_j,\zeta)} \right\}. \tag{2.4}$$

(23) and (26) have shown that the maximum likelihood estimator of d is strongly consistent, asymptotically normally distributed and asymptotically efficient in the Fisher sense.

A Monte Carlo study analyzing the behaviour of the finite sample efficiency of the maximum likelihood estimators using an approximate frequency-domain (4) and the exact time-domain (20) approaches may be found in 27). These studies indicate that for an ARFIMA $(0,d,0)$ model with unknown mean the results are very similar. (2) also evaluate the performance of Sowell's method (20) with the approximate Gaussian maximum likelihood procedure suggested in (7).

3. Identification and Estimation of an ARFIMA(p,d,q) Model

For the use of the regression techniques several steps are necessary to obtain an ARFIMA model for a set of time series data and these are given below (see (3) and (28)).

Let $\{X_t\}$ be the process as defined in (2.1). Then $U_t = (1-B)^d X_t$ is an ARMA(p,q) process and $Y_t = \frac{\Phi(B)}{\Theta(B)} X_t$ is an ARFIMA$(0,d,0)$ process.

Model Building Steps:

1. Estimate d in the ARIMA(p,d,q) model; denote the estimate by $\hat{d}$.

2. Calculate $\hat{U}_t = (1-B)^{\hat{d}} X_t$.

3. Using Box-Jenkins modelling procedure (see (29)) (or the AIC criterion, (30)) identify and estimate ϕ and θ parameters in the ARMA(p,q) process $\phi(B)\hat{U}_t = \theta(B)\epsilon_t$.

4. Calculate $\hat{Y}_t = \frac{\hat{\phi}(B)}{\hat{\theta}(B)} X_t$.

5. Estimate d in the ARFIMA$(0,d,0)$ model $(1-B)^{\hat{d}}\hat{Y}_t = \epsilon_t$. The value of $\hat{d}$ obtained in this step is now the new estimate of d.

6. Repeat steps 2 to 5, until the estimates of the parameters d, ϕ and θ converge.

In this algorithm, to estimate d we use the regression methods described in Section 2. It should be noted that usually only one iteration with Steps 1-3 is used to obtain a model (see, for instance, (28)). Related to Step 3, it has widely been discussed that the bias in the estimator of d can lead to the problem of identifying the short-memory parameters. This issue has been investigated by (31), (32) and recently, by (1) and (33).

4. Simulation Study

Now we investigate, by simulation experiments, the convergence of the iterative method of model estimation shown in Section 3. In this study, observations from the ARFIMA(p, d, q) process are generated using the method described in (34) where the random variables ϵ_t are assumed to be identically and independently normally distributed as $N(0, 1.0)$ obtained from the subroutine RNNOR in the IMSL - Library. For the estimators $\hat{d}_p$ and $\hat{d}_{sp}$, we use $g(n) = n^{0.5}$ and $m = n^{0.9}$ (the truncation point in the Parzen lag window), as suggested in (21) and (9), respectively. In the case of Robinson's estimator we use $l = 2$, $\tau = 0.5$ and $A(d, \tau) = 1.0$. The respective numbers of observations involved in the regression equations are given in the tables. Three models are considered: ARFIMA$(0, d, 0)$, ARFIMA$(1, d, 0)$ and ARFIMA$(0, d, 1)$. ARFIMA $(0, d, 0)$ model is included here to verify the finite sample behaviour and also the performance of the smoothed periodogram function in the Robinson's method.

In the Whittle method, the parameters of the process are estimated simultaneously by the subroutine BCONF in the IMSL - Library. In the case of the semiparametric methods, the autoregressive and moving average parameters are estimated by the subroutine NSLE in the IMSL - Library, after the time series has been differentiated by the estimate of d. In our simulation, we assume that the true model is known and only the parameters need to be estimated. The results for all estimation procedures are based on the same 500 replications.

ARFIMA$(0, d, 0)$:

Table 4.1 gives the mean value of $\hat{d}$ (mean $(\hat{d})$), the standard deviation (sd, in parenthesis), the bias $(\hat{d})$, the mean squared error (mse), and the values of $g(n)$ (the upper limit of the frequencies involved in the semiparametric approaches). As expected, the Whittle's method for estimating d is more accurate than the other methods. Nevertheless, the other methods give good results as well. The results get better when the sample size increases. For the Robinson methods, the choice of the number of frequencies is crucial for estimating d. For $d = 0.2$, $\hat{d}_{pr}$ and $\hat{d}_{spr}$ have bigger mean squared errors compared to the other methods. In this case, the regression is built from $l = 2, \cdots, g(n)$, that is, less observations are used to obtain $\hat{d}_{pr}$ and $\hat{d}_{spr}$. For $d = 0.45$, both estimators improve with smaller bias and mean squared error and they are very competitive to the Whittle's estimator. $\hat{d}_{spr}$ dominates $\hat{d}_{pr}$ and $\hat{d}_{sp}$ outperform $\hat{d}_p$ in terms of mean squared error.

Table 4.1: Estimation of d: ARFIMA (0,d,0)

n	d	$\hat{d}_W$	$\hat{d}_{sp}$	$\hat{d}_p$	$\hat{d}_{spr}$	$\hat{d}_{pr}$
150	0.2					
	mean($\hat{d}$)	0.1983	0.1396	0.2110	0.2153	0.2252
	sd	(0.0749)	(0.1915)	(0.2470)	(0.2862)	(0.4289)
	bias ($\hat{d}$)	-0.0017	-0.0604	0.0110	0.0153	0.0252
	mse	0.0056	0.0402	0.0610	0.0819	0.1841
	$g(n)$				12	12
	0.45					
	mean($\hat{d}$)	0.4768	0.3724	0.4500	0.4653	0.4615
	sd	(0.0379)	(0.1879)	(0.2275)	(0.0828)	(0.1108)
	bias	0.0268	-0.0776	0.0	0.0153	0.0115
	mse	0.0021	0.0412	0.0516	0.0071	0.0124
	$g(n)$				65	65
300	0.2					
	mean($\hat{d}$)	0.2033	0.1562	0.2018	0.2175	0.2075
	sd	(0.0494)	(0.1501)	(0.1970)	(0.2160)	(0.3088)
	bias	0.0033	-0.0438	0.0018	0.0175	0.0075
	mse	0.0024	0.0244	0.0387	0.0468	0.0952
	$g(n)$				17	17
	0.45					
	mean($\hat{d}$)	0.4721	0.4020	0.4594	0.4593	0.4556
	sd	(0.0351)	(0.1631)	(0.2040)	(0.0646)	(0.0835)
	bias	0.0221	-0.0480	0.0094	0.0093	0.0056
	mse	0.0017	0.0218	0.0416	0.0043	0.007
	$g(n)$				115	115

It should be noted that $n = 300$ may not be large enough for some of the methods to perform better. To get a feel about the asymptotic behaviour of these methods we conducted a study with $n = 10,000, d = .2$ and one replication. The results are in Table 4.2. The case $n = 300$ is also given for comparison. The bias of all methods decrease substantially when $n = 10,000$ with $\hat{d}_w$ having the smallest bias.

Table 4.2: Asymptotic performance of $\hat{d}$: ARFIMA (0,d,0)

(One replication only)

n	d	$\hat{d}_W$	$\hat{d}_{sp}$	$\hat{d}_p$	$\hat{d}_{spr}$	$\hat{d}_{pr}$
300	0.2					
	$\hat{d}$	0.29366	0.40769	0.52379	0.42221	0.44927
	bias	0.09366	0.20769	0.32379	0.22221	0.24927
	$g(n)$		17	17	17	17
10,000	0.2					
	$\hat{d}$	0.19652	0.17648	0.16541	0.18818	0.15482
	bias	-0.00348	-0.02352	-0.03459	-0.01182	-0.04518
	$g(n)$		100	100	100	100
300	0.45					
	$\hat{d}$	0.56356	0.62544	0.76787	0.58141	0.53576
	bias	0.11356	0.17544	0.31787	0.13141	0.08576
	$g(n)$		17	17	115	115
10,000	0.45					
	$\hat{d}$	0.45213	0.42407	0.45848	0.43736	0.44313
	bias	0.00213	-0.02593	0.00848	-0.01264	-0.00687
	$g(n)$		100	100	2154	2154

ARFIMA (p, d, q) MODELS:

These models contain short memory components and the estimation of all parameters is the goal. Thus, the long memory parameter d is estimated taking into account the additional uncertainty due to the contemporary estimation of the autoregressive or moving average parameters.

Following the procedure described in Section 3, for each d, ϕ and θ we generate a time series of size $n = 300$, estimate the fractional parameter d and then obtain $\hat{U}_t = (1 - B)^{\hat{d}} X_t$ (see Step 2 in Section 3) from which the autoregressive or the moving average coefficient estimate is obtained as in

Step 3. Then we obtain $\hat{Y}_t = (\hat{\phi}(B)/\hat{\theta}(B))X_t$ which is an ARFIMA$(0,d,0)$ process and use it to estimate d. Steps 2-5 are repeated until the values of $(\hat{d}, \hat{\phi}, \hat{\theta})$ do not change much from one iteration to the next. In each iteration d is estimated using $\hat{d}_p$, $\hat{d}_{sp}$, $\hat{d}_{pr}$ and $\hat{d}_{spr}$. This procedure is repeated 500 times. In each replication, the maximum number of iterations is fixed at 20. An extensive simulation study was performed considering different values of d, ϕ and θ with $p = q = 0, 1$. However, we only present some of them here since the pattern is the same for the other cases.

The results are shown in Tables 4.3 to 4.8. The first part of the tables (I) gives the results corresponding to the first iteration. These are the average of $\hat{d}$, (mean($\hat{d}$)), bias, sd, mean squared error (mse), the average of the coefficient estimate (mean ($\hat{\phi}$) or mean ($\hat{\theta}$)), bias in the coefficient estimate and the sd of the coefficient obtained from the first iteration over the 500 replications. The second part of the tables (II) gives the value of l_i, the maximum iteration to obtain convergence, and the corresponding estimation results as in part I. Note that, in the second part of the table, there are no results for the Whittle's method. In the tables, the smallest values of bias and mse are in bold face.

From the results we can discuss the following issues:

i. The number of iterations (l_i) needed to obtain convergence for the estimates.

ii. The impact of the values of d, ϕ, θ for convergence. The convergence of the parameter estimates to the true values.

iii. The behaviour of the estimators $\hat{d}_p$, $\hat{d}_{sp}$, $\hat{d}_{pr}$, $\hat{d}_{spr}$ and $\hat{d}_W$.

iv. The comparison between parametric and semiparametric methods.

ARFIMA $(1, d, 0)$:

Table 4.3: Estimation for d = 0.2: ARFIMA(1,d,0), ϕ = -0.2

	$d = 0.2$			$\phi =$ -0.2		
	i	p	sp $g(n) = 17$	pr	spr $g(n) = 17$	w
I	mean($\hat{d}_i$)	0.2507	0.1950	0.2511	0.2450	0.1902
	bias($\hat{d}_i$)	0.0507	**-0.0050**	0.0511	0.0450	-0.0098
	sd($\hat{d}_i$)	0.1269	0.0734	0.2094	0.1103	0.0734
	mse($\hat{d}_i$)	0.0186	**0.0054**	0.0464	0.0142	0.0055
	mean($\hat{\phi}_i$)	-0.2245	-0.1875	-0.1935	-0.2232	-0.1913
	bias($\hat{\phi}_i$)	-0.0245	0.0125	**0.0065**	-0.0232	0.0087
	sd($\hat{\phi}_i$)	0.1233	0.0917	0.2342	0.1156	0.0911
II	l_i	2	2	7	2	–
	mean($\hat{d}_i^*$)	0.2534	0.1977	0.2386	0.2490	–
	bias($\hat{d}_i^*$)	0.0534	**-0.0023**	0.0386	0.0490	–
	sd($\hat{d}_i^*$)	0.1290	0.0743	0.2695	0.1119	–
	mean($\hat{\phi}_i^*$)	-0.2262	-0.1898	-0.1856	-0.2261	–
	bias($\hat{\phi}_i^*$)	-0.0262	**0.0102**	0.0144	-0.0261	–
	sd($\hat{\phi}_i^*$)	0.1252	0.0924	0.2692	0.1170	–

Table 4.4: **Estimation for $d = 0.2$: ARFIMA(1,d,0), $\phi = 0.2$**

	$d = 0.2$		$\phi = 0.2$			
	i	p	sp $g(n) = 17$	pr	spr $g(n) = 17$	w
I	mean($\hat{d}_i$)	0.2568	0.1942	0.2610	0.2428	0.1762
	bias($\hat{d}_i$)	0.0568	**-0.0058**	0.0610	0.0428	-0.0238
	sd($\hat{d}_i$)	0.1268	0.0683	0.1984	0.1034	0.1295
	mse($\hat{d}_i$)	0.0193	**0.0047**	0.0430	0.0125	0.0173
	mean($\hat{\phi}_i$)	0.1530	0.2093	0.1633	0.1623	0.2177
	bias($\hat{\phi}_i$)	-0.0470	**0.0093**	-0.0367	-0.0377	0.0177
	sd($\hat{\phi}_i$)	0.1384	0.0941	0.2126	0.1206	0.1394
II	l_i	6	3	6	6	–
	mean($\hat{d}_i^*$)	0.2496	0.1854	0.2103	0.2329	–
	bias($\hat{d}_i^*$)	0.0496	-0.0146	**0.0103**	0.0329	–
	sd($\hat{d}_i^*$)	0.1340	0.0729	0.3221	0.1131	–
	mean($\hat{\phi}_i^*$)	0.1615	0.2194	0.1965	0.1739	–
	bias($\hat{\phi}_i^*$)	-0.0385	0.0194	**-0.0035**	-0.0261	–
	sd($\hat{\phi}_i^*$)	0.1484	0.1017	0.2759	0.1354	–

Table 4.5: Estimation for d = 0.45: ARFIMA(1,d,0), ϕ = -0.2

	$d = 0.45$		ϕ = -0.2			
	i	p	sp $g(n) = 17$	pr	spr $g(n) = 115$	w
I	mean($\hat{d}_i$)	0.5123	0.4449	0.3616	0.3679	0.5230
	bias($\hat{d}_i$)	0.0623	**-0.0051**	-0.0884	-0.0821	0.0730
	sd($\hat{d}_i$)	0.1296	0.0739	0.0747	0.0563	0.0800
	mse($\hat{d}_i$)	0.0206	**0.0055**	0.0134	0.0099	0.0117
	mean($\hat{\phi}_i$)	-0.2234	-0.1773	-0.0848	-0.0981	-0.2568
	bias($\hat{\phi}_i$)	-0.0234	**0.0227**	0.1152	0.1019	-0.0568
	sd($\hat{\phi}_i$)	0.1389	0.0981	0.0993	0.0711	0.0815
II	l_i	5	3	9	4	–
	mean($\hat{d}_i^*$)	0.5154	0.4475	0.3308	0.4459	–
	bias($\hat{d}_i^*$)	0.0654	**-0.0025**	-0.1192	-0.0041	–
	sd($\hat{d}_i^*$)	0.1310	0.0750	0.3850	0.1333	–
	mean($\hat{\phi}_i^*$)	-0.2254	-0.1796	-0.0446	-0.1695	–
	bias($\hat{\phi}_i^*$)	-0.0254	**0.0204**	0.1554	0.0305	–
	sd($\hat{\phi}_i^*$)	0.1410	0.0997	0.4273	0.1960	–

Table 4.6: Estimation for d = 0.45: ARFIMA(1,d,0), ϕ = 0.2

	$d = 0.45$		$\phi = 0.2$			
	i	p	sp $g(n) = 17$	pr	spr $g(n) = 115$	w
I	mean($\hat{d}_i$)	0.5097	0.4491	0.5928	0.5958	0.6362
	bias($\hat{d}_i$)	0.0597	**-0.0009**	0.1428	0.1458	0.1862
	sd($\hat{d}_i$)	0.1231	0.0725	0.0741	0.0579	0.1471
	mse($\hat{d}_i$)	0.0187	**0.0052**	0.0259	0.0246	0.0562
	mean($\hat{\phi}_i$)	0.1552	0.2139	0.0642	0.0601	0.0376
	bias($\hat{\phi}_i$)	-0.0448	**0.0139**	-0.1358	-0.1399	-0.1624
	sd($\hat{\phi}_i$)	0.1426	0.1005	0.0654	0.0503	0.1322
II	l_i	8	6	10	10	–
	mean($\hat{d}_i^*$)	0.5009	0.4393	0.3581	0.4118	–
	bias($\hat{d}_i^*$)	0.0509	**-0.0107**	-0.0919	-0.0382	–
	sd($\hat{d}_i^*$)	0.1318	0.0781	0.3690	0.2923	–
	mean($\hat{\phi}_i^*$)	0.1664	0.2266	0.3026	0.2500	–
	bias($\hat{\phi}_i^*$)	-0.0336	**0.0266**	0.1026	0.0500	–
	sd($\hat{\phi}_i^*$)	0.1573	0.1123	0.3627	0.3062	–

Tables 4.3 to 4.6 present the results corresponding to $d = 0.2, 0.45$ and $\phi = -0.2, 0.2$. We summarize the findings as follows:

i. The number of iterations to stabilize the estimates increases with ϕ and d, and its value is larger when ϕ is positive. In most of the cases considered here, the estimates of d and ϕ obtained in the first iteration (steps 1-3) and the converged ones are very close. The computational effort involved in the iterative procedure is not simple and the problem of order identification when the time series is differenciated many times must also be considered. In certain cases there were difficulties to achieve convergence of the parameters, especially for those closer to the non-stationary boundary in the Robinson's method. Thus, we feel that only one iteration (steps 1-3) is needed in the model building algorithm described in Section 3. We also computed the averages of the standard deviations calculated from the estimates in the 20 iterations in each replication (the results are not presented here). These values are very small and they indicate that the changes in the values of the estimates from iteration to iteration are very small. This confirms our earlier assertion that estimates from the first iteration would be sufficient for practical purposes.

ii. The estimation of AR coefficients do impact the estimation of d and also the iterative procedure in section 3. When $\phi > 0$ biases of all the estimators of d increase with ϕ. When $|\phi|$ is large the biases in all estimators of d are large (except for $\hat{d}_{sp}$), so are the biases in the estimators of ϕ but in the opposite direction. This indicates that the bias in $\hat{d}$ is being compensated by the bias in $\hat{\phi}$. When ϕ is negative the estimates of the parameters are typically better behaved than in the positive case. Also, the number of iterations needed to attain convergence is smaller (compare, for instance, the cases ARFIMA$(1, 0.45, 0)$ when $\phi = -0.2$ and $\phi = 0.2$).

iii. $\hat{d}_{sp}$ has smaller mean squared error and, in general, also has smaller bias compared to the other regression estimators. When ϕ is negative, $\hat{d}_{sp}$ underestimates d while $\hat{d}_p$, $\hat{d}_{pr}$ and $\hat{d}_{spr}$ overestimates d most of the time except when, $d = 0.45$ where $\hat{d}_{pr}$ and $\hat{d}_{spr}$ underestimate the true value. $\hat{d}_{sp}$, and its corresponding $\hat{\phi}$, move more rapidly to true values compared to $\hat{d}_p$, and its corresponding $\hat{\phi}$. Also, as expected, s.d.$(\hat{d}_{sp}) <$ s.d.$(\hat{d}_p)$. It should also be noted that the simulated standard deviations are close to the asymptotic values. For instance, when $d = 0.45$ and $\phi = 0.2$ the simulated standard deviations for $\hat{d}_{sp}$ and $\hat{d}_p$ are 0.0725 and 0.1231, respectively, while the asymptotic values are 0.0876 and 0.2018, respectively. It is clear that the biases of $\hat{d}_{pr}$ and $\hat{d}_{spr}$ are more pronounced than those of the usual $\hat{d}_p$ and $\hat{d}_{sp}$ estimators. The first two methods involve more frequencies in the regression equation and this yields estimates with large bias and mean squared error, especially when ϕ is positive and d is large. This may be caused by the fact that the AR component enlarges the value of the spectral density function. The results are different from the ones in the ARFIMA$(0, d, 0)$ model. $\hat{d}_{spr}$ has a smaller mean squared error compared with $\hat{d}_{pr}$ as expected since the spectral density function is estimated by the smoothed periodogram function.

iv. For large and positive ϕ, the semiparametric methods, especially the smoothed periodogram performs better than the Whittle's method which improves when ϕ is negative and not closer to the non-stationary boundary.

ARFIMA$(0, d, 1)$:

Table 4.7: Estimation for d = 0.3: ARFIMA(0,d,1), θ = -0.3

	d = 0.3	θ = -0.3				
	i	p	sp $g(n) = 17$	pr	spr $g(n) = 23$	w
I	mean($\hat{d}_i$)	0.3458	0.2962	0.3528	0.3501	0.3153
	bias($\hat{d}_i$)	0.0458	**-0.0038**	0.0528	0.0501	0.0153
	sd($\hat{d}_i$)	0.1315	0.0761	0.1967	0.1234	0.0059
	mse($\hat{d}_i$)	0.0193	0.0058	0.0413	0.0177	**0.0037**
	mean($\hat{\theta}_i$)	-0.2624	-0.3046	-0.2519	-0.2571	-0.2897
	bias($\hat{\theta}_i$)	0.0376	**-0.0046**	0.0481	0.0429	0.0103
	sd($\hat{\theta}_i$)	0.1270	0.0903	0.1844	0.1262	0.0763
II	l_i	3	3	15	3	–
	mean($\hat{d}_i^*$)	0.3423	0.2922	0.3466	0.3427	–
	bias($\hat{d}_i^*$)	0.0423	**-0.0078**	0.0466	0.0427	–
	sd($\hat{d}_i^*$)	0.1324	0.0767	0.2044	0.1258	–
	mean($\hat{\theta}_i^*$)	-0.2652	-0.3078	-0.2559	-0.2632	–
	bias($\hat{\theta}_i^*$)	0.0348	**-0.0078**	0.0441	0.0368	–
	sd($\hat{\theta}_i^*$)	0.1277	0.0907	0.1957	0.1282	–

Table 4.8: Estimation for $d = 0.3$: ARFIMA(0,d,1), $\theta = 0.3$

		$d = 0.3$		$\theta = 0.3$		
	i	p	sp $g(n) = 17$	pr	spr $g(n) = 23$	w
I	mean($\hat{d}_i$)	0.3409	0.2788	0.2581	0.2804	0.3385
	bias($\hat{d}_i$)	0.0409	**-0.0212**	-0.0419	-0.0196	0.0385
	sd($\hat{d}_i$)	0.0999	0.0686	0.1544	0.1005	0.1018
	mse($\hat{d}_i$)	0.0116	**0.0051**	0.0255	0.0104	0.0118
	mean($\hat{\theta}_i$)	0.3365	0.2703	0.2422	0.2704	0.3288
	bias($\hat{\theta}_i$)	0.0365	**-0.0297**	-0.0578	-0.0296	0.0288
	sd($\hat{\theta}_i$)	0.1242	0.0898	0.1801	0.1168	0.1154
II	l_i	10	6	15	15	–
	mean($\hat{d}^*_i$)	0.3638	0.2924	0.2989	0.3186	–
	bias($\hat{d}^*_i$)	0.0638	**-0.0076**	-0.0011	0.0186	–
	sd($\hat{d}^*_i$)	0.1149	0.0755	0.1981	0.1337	–
	mean($\hat{\theta}^*_i$)	0.3607	0.2851	0.2826	0.3097	–
	bias($\hat{\theta}^*_i$)	0.0607	-0.0149	-0.0174	**0.0097**	–
	sd($\hat{\theta}^*_i$)	0.1410	0.0987	0.2171	0.1485	–

Simulation results for the ARFIMA$(0, d, 1)$ process are given in Tables 4.6 to 4.7. Although we considered several values of d and θ we present the results only for $d = 0.3$ and $\theta = -0.3, 0.3$ to save space. We note that more iterations are needed when $\theta > 0$. The estimator $\hat{d}_{sp}$ outperforms the other methods including the Whittle's estimator $\hat{d}_W$. The biases in $\hat{d}_{sp}$ and $\hat{\theta}$ increases when θ is positive.

As in the ARFIMA$(1, d, 0)$ model, $\hat{d}_p$ and $\hat{d}_{sp}$ need only small number of iterations to achieve convergence with the latter requiring the smallest. Results for the estimators $\hat{d}_{pr}$ and $\hat{d}_{spr}$ are not very good. If we consider only one iteration then, in general, the two regression estimators perform much better than $\hat{d}_{pr}$ and $\hat{d}_{spr}$.

We also encountered some convergence difficulties for the Robinson's estimator $\hat{d}_{pr}$ especially for positive and large values of θ. In most of the cases, the least squares estimation of the parameters failed to converge. Both $\hat{d}_{pr}$ and $\hat{d}_{spr}$ estimators, have very large sample variances. Extensive computational efforts were necessary to obtain 500 successful replications with a maximum of 20 iterations in each.

5. Summary and Concluding Remarks

In this paper we considered a simulation study to evaluate the procedures for estimating the parameters of an ARFIMA process. We considered both parametric and semiparametric methods and also used the smoothed periodogram function in the modified regression estimator. The results indicate that the regression methods outperforms the parametric Whittle's method when AR or MA components are involved. Performance of the Robinson estimator usually is not as good as the other semiparametric methods; it has large bias, standard deviation, and mean squared error. The use of the smoothed periodogram in Robinson's method improves the estimates, however, the results are still not as good as the usual regression methods. The results also indicate that the estimates from the first iteration (steps 1-3) are sufficient for practical purposes.

Acknowledgements

B. Abraham was partially supported by a grant from NSERC. V.A. Reisen was partially supported by CNPq-Brazil. S. Lopes was partially supported by Pronex *Fenômenos Críticos em Probabilidade e Processos Estocásticos* (Convênio FINEP/MCT/CNPq 41/ 96/0923/00) and CNPq-Brazil. We would like to thank Dr. Ela Mercedes Toscano (UFMG - BRAZIL) for some help with the simulations and Dr. Bonnie K. Ray (NJIT - USA) for providing the computer code for the Whittle's estimator. We also thank the anonymous referees for their helpful comments.

Bibliography

(1) Smith, J., Taylor, N. and Yadav, S., "Comparing the Bias and Misspecification in ARFIMA Models". *Journal of Time Series Analysis*, 1997, 18(5), 507-527.

(2) Beveridge, S. and Oickle, C., "Estimating Fractionally Integrated Time Series Models". *Economics Letters*,1993, 43, 137-142.

(3) Hosking, J., "Fractional Differencing". *Biometrika*,1981, 68(1), 165-176.

(4) Fox, R. and Taqqu, M.S., "Large-sample Properties of Parameter Estimates for Strongly Dependent Stationary Gaussian Time Series". *The Annals of Statistics*, 1986, 14(2), 517-532.

(5) Robinson, P.M., "Log-Periodogram Regression of Time Series with Long

Range Dependence". *The Annals of Statistics*,1995a, 23(3), 1048-1072.

(6) Beran, J., *Statistics for Long Memory Processes.* New York: Chapman and Hall, 1994

(7) Li, W.K. and McLeod, A.I., "Fractional Time Series Modelling". *Biometrika*, 1986, 73(1), 217-221.

(8) Hassler, U., "Regression of Spectral Estimator with Fractionally Integrated Time Series". *Journal of Time Series Analysis*, 1993, 14, 369-380.

(9) Reisen, V.A., "Estimation of the Fractional Difference Parameter in the ARIMA(p, d, q) Model Using the Smoothed Periodogram". *Journal of Time Series Analysis*, 1994, 15(3), 335-350.

(10) Chen, G., Abraham, B. and Peiris, S., "Lag Window Estimation of the Degree of Differencing in Fractionally Integrated Time Series Models". *Journal of Time Series Analysis*, 1994, 15(5), 473-487.

(11) Robinson, P.M., "Gaussian Semiparametric Estimation of Long Range Dependence". *The Annals of Statistics*, 1995b, 23(5), 1630-1661.

(12) Taqqu, M.S., Teverovsky, V. and Bellcore, W. W., "Estimators for Long-Range Dependence: an Empirical Study". *Fractals*, 1995, 3(4), 785-802.

(13) Taqqu, M. S. and Teverovsky, V., "Robustness of Whittle-type Estimators for Time Series with Long-Range Dependence". *Technical Report*, Boston University, Massachussetts, 1996.

(14) Lobato, I. and Robinson, P. M., "Averaged Periodogram Estimation of

Long Memory", *Journal of Econometrics*, 1996, 73, 303-324.

(15) Bisaglia, L. and Guégan, D., "A Comparison of Techniques of Estimation in Long Memory Processes". *Computational Statistics and Data Analysis*, 1998, 27, 61-81.

(16) Hurvich, C. M., Deo, R. S. and Brodsky, J., "The Mean Squared Error of Geweke and Porter-Hudak's Estimator of the Memory Parameter of a Long Memory Time Series". *Journal of Time Series Analysis*, 1998, 19(1), 19-46.

(17) Hurvich, C.M. and Deo, R.S., "Plug-in Selection of the Number of Frequencies in Regression Estimates of the Memory Parameter of a Long-Memory Time Series ". *Journal of Time Series Analysis*, 1999, 20(3), 331-341.

(18) Velasco, C., "Gaussian Semiparametric Estimation of Non-stationary Time Series". *Journal of Time Series Analysis*, 1999, 20(1), 87-127.

(19) Chong, T., T., "Estimating the Differencing Parameter via the Partial Autocorrelation Function". *Journal of Econometrics*, 2000, 97, 365-381.

(20) Sowell, F., "Maximum Likelihood Estimation of Stationary Univariate Fractionally Integrated Time Series Models". *Journal of Econometrics*, 1992, 53, 165-188.

(21) Geweke, J. and Porter-Hudak, S., "The Estimation and Application of Long Memory Time Series Models". *Journal of Time Series Analysis*, 1983, 4(4), 221-238.

(22) Giraitis, L. and Surgailis, D., "A Central Limit Theorem for Quadratic Forms in Strongly Dependent Linear Variables and its Application to Asymptotical Normality of Whittle's Estimate", *Probability Theory and Related Fields*, 1990, 86, 87-104.

(23) Dahlhaus, R., "Efficient Parameter Estimation for Self-Similar Processes". *The Annals of Statistics*, 1989, 17(4), 1749-1766.

(24) Robinson, P. M., "Rates of Convergence and Optimal Spectral Bandwidth for Long Range Dependence", *Probability and Theory Related Fields*, 1994, 99, 443-473.

(25) Whittle, P., "Estimation and Information in Stationary Time Series". *Arkiv för Matematik*, 1953, 2, 423-434.

(26) Yajima, Y., "On Estimation of Long Memory Time Series Models". *The Australian Journal of Statistics*, 1985, 27, 303-320.

(27) Cheung, Y. and Diebold, F.X., "On Maximum Likelihood Estimation of the Differencing Parameter of Fractionally-Integrated Noise with Unknown Mean". *Journal of Econometrics*, 1994, 62, 301-316.

(28) Brockwell, P.J. and Davis, R.A., *Time Series: Theory and Methods*. Springer-Verlag: New York, 1991.

(29) Box, G.E.P. and Jenkins, G.M., *Time Series Analysis; Forecasting and Control*. 2nd ed., Holden-Day: San Francisco, 1976.

(30) Akaike, H., "Maximum Likelihood Identification of Gaussian Autore-

gressive Moving Average Models". *Biometrika*,1973, 60(2), 255-265.

(31) Schmidt, C. M. and Tschernig, R., "Identification of Fractional ARIMA Models in the Presence of Long Memory". *Paper presented at the FAC Workshop on Economic Time Series Analysis and System Identification*, July, Vienna, 1993.

(32) Crato, N. and Ray, B.K., "Model Selection and Forecasting for Long-range Dependent Processes". *Journal of Forecasting*, 1996, 15, 107-125.

(33) Reisen, V.A. and Lopes, S., "Some Simulations and Applications of Forecasting Long-Memory Time Series Models". *Journal of Statistical Planning and Inference*, 1999, 80(2), 269-287.

(34) Hosking, J., "Modelling Persistence in Hydrological Time Series using Fractional Differencing". *Water Resources Research*, 1984, 20(12), 1898-1908.