A Pseudo-Linear Method for Fractionally Integrated Arma (ARFIMA) Model Estimation

G.N. Fouskitakis and *S.D. Fassois*
Department of Mechanical Engineering
University of Patras
GR-265 00 Patras, Greece

Abstract

A Pseudo-Linear method for the estimation of Fractionally Integrated ARMA (ARFIMA) models is introduced. The method uses a long binomial series expansion of the fractional differencing operator, as well as the relationship of the AR/MA parameters and binomial expansion terms with the model's inverse function. It is based upon a pseudo-linear formulation motivated by the fact that this relationship leads to a special-form regression problem that can be decomposed into a scalar non-linear and a multiple linear regression. The performance characteristics of the method are demonstrated via Monte Carlo experiments and comparisons with the frequency-domain Maximum Likelihood method.

1. Introduction

Most of the work on time series analysis has been concerned with series characterized by the property that distant observations behave independently, or nearly so. Yet, in many empirical studies [1-3] the dependence between distant observations is not negligible and decays very slowly. Series with such long-term persistence are referred to as *long-memory* time series, and their power spectral density increases indefinitely as the frequency approaches zero, while their autocorrelation decays hyperbolically.

Long-memory time series aren't well represented by the usual stationary AutoRegressive Moving Average ($ARMA$) models, which are characterized by limited, at the origin, power spectral density and an exponentially decaying autocorrelation [4].

A class of models that exhibits the foregoing long-memory characteristics is that of *Fractionally Integrated ARMA (ARFIMA)* models. This in essence is an extension of the Integrated ARMA (ARIMA) models of Box *et al.* [4], in which the differencing operator is raised into a fractional, instead of the usual integer, power.

The majority of the available $ARFIMA$ model estimation methods follow a two-step approach, according to which an estimate of the fractional power is obtained (usually in the frequency domain) in the first step, and a standard $ARMA$ estimation technique is applied to the adjusted (filtered by the fractional differencing operator) time series in the second. These methods have been criticized for failing to produce good estimates for relatively short data records [2]. The alternative one-step methods advocate the simultaneous estimation of all model parameters based upon variants of the Maximum Likelihood procedure in either the time or the frequency domains [2,5-7]. A major drawback of this category of methods is their high computational complexity.

In this paper a simple and computationally efficient Pseudo-Linear method for $ARFIMA$ model estimation is introduced. The method uses exclusively time-domain operations and is based upon the decomposition of a special-form regression problem into a scalar non-linear and a multiple linear regression.

2. Problem statement

An $ARFIMA(n,d,m)$ process is of the form:

$$\Phi(B)\cdot(1-B)^d\cdot X_t = \Theta(B)\cdot\alpha_t$$
$$\alpha_t \sim i.i.d.\mathcal{N}(0,\sigma_\alpha^2) \quad d\in(-0.5,0.5) \qquad (1)$$

with t indicating discrete time, X_t the observed time series, α_t an independently identically distributed ($i.i.d.$) Gaussian sequence with the indicated mean and variance, B the backshift operator ($BX_t \triangleq X_{t-1}$), d the fractional power, and $\Phi(B)$, $\Theta(B)$ the autoregressive

Proceedings of the 8th IEEE Signal Processing Workshop on Statistical Signal and Array Processing
Corfu, Greece, 1996

(AR) and moving average (MA) polynomials:

$$\Phi(B) \triangleq 1 + \sum_{i=1}^{n} \phi_i B^i \qquad \Theta(B) \triangleq 1 + \sum_{i=1}^{m} \theta_i B^i \quad (2)$$

Fractional differencing is defined by the binomial series expansion:

$$(1-B)^d = 1 + \sum_{j=1}^{\infty} \alpha_j \cdot B^j = 1 + \alpha_1 \cdot B + \alpha_2 \cdot B^2 + \ldots \quad (3)$$

$$\alpha_1 = -d, \quad \alpha_j = \frac{\alpha_{j-1} \cdot (j-1-d)}{j} \quad (j = 2, 3, \ldots) \quad (4)$$

The process representation (1) is assumed to satisfy the following standard assumptions:

A1. $d \in \mathcal{D} \triangleq (-0.5, 0.5)$, $\Phi(B) \neq 0$ for $|B| \leq 1$ *(stationarity conditions)*

A2. $\Theta(B) \neq 0$ for $|B| \leq 1$ *(invertibility condition)*

The problem of *ARFIMA* process estimation may be then stated as follows: "Given time series data X_t $t \in [1, N]$, select a particular model $\mathcal{M}(\bar{\mathbf{p}})$ from the model set[1]:

$$\mathcal{M} = \{\mathcal{M}(\bar{\mathbf{p}}) : \Phi(B) \cdot (1-B)^d \cdot X_t = \Theta(B) \cdot e_t(\bar{\mathbf{p}})$$
$$\bar{\mathbf{p}} \triangleq [d\ \boldsymbol{\phi}^T\ \boldsymbol{\theta}^T\ \sigma_e^2]^T \in \mathcal{D} \times \mathcal{H}(\Phi) \times \mathcal{H}(\Theta) \times \Re^+\} \quad (5)$$

where $e_t(\bar{\mathbf{p}})$ represents the model's one-step-ahead prediction error, σ_e^2 its variance, $\boldsymbol{\phi}$, $\boldsymbol{\theta}$ the *AR* and *MA* parameter vectors, respectively, and $\mathcal{H}(\Phi)$, $\mathcal{H}(\Theta)$ the regions of $\Re^n$, $\Re^m$ in which the stationarity and invertibility, respectively, conditions hold."

3. The pseudo-linear estimation method

The substitution of a truncated, p-th order, binomial series expansion of the fractional differencing operator (3) into the *ARFIMA* representation (1) yields the $ARMA(p+n, m)$ representation:

$$(1 + \rho_1 \cdot B + \ldots + \rho_{p+n} \cdot B^{p+n}) \cdot X_t = (1 + \theta_1 \cdot B + \ldots \theta_m \cdot B^m) \cdot \alpha_t \quad (6)$$

with ρ_i defined by the convolution expressions:

$$\rho_i = \sum_{k=0}^{i} \alpha_k \phi_{i-k} \qquad (i = 1, 2, \ldots, p+n) \quad (7)$$

[1]Bold face lower-case/capital characters represent vector/matrix quantities.

with $\alpha_0 \equiv 1$, $\alpha_k \equiv 0$ $(k > p)$, $\phi_0 \equiv 1$, $\phi_k \equiv 0$ $(k > n)$. Denoting this representation's inverse function operator as:

$$I(B) \triangleq 1 + \sum_{i=1}^{\infty} I_i B^i \triangleq P(B)/\Theta(B) \quad (8)$$

and combining it with (7) yields:

$$\begin{aligned}
&\alpha_1 + \Phi_1 - \Theta_1 = I_1 \\
&\alpha_2 + \alpha_1\Phi_1 + \Phi_2 - I_1\Theta_1 - \Theta_2 = I_2 \\
&\qquad\qquad \vdots \qquad\qquad \vdots \\
&\alpha_p + \alpha_{p-1} \cdot \Phi_1 + \ldots + \alpha_{p-n} \cdot \Phi_n - I_{p-1} \cdot \Theta_1 - \ldots \\
&-I_{p-m} \cdot \Theta_m = I_p \\
&\alpha_p \cdot \Phi_1 + \ldots + \alpha_{p-n+1} \cdot \Phi_n - I_p \cdot \Theta_1 - \ldots \\
&-I_{p-m+1} \cdot \Theta_m = I_{p+1} \\
&\alpha_p \cdot \Phi_2 + \ldots + \alpha_{p-n+2} \cdot \Phi_n - I_{p+1} \cdot \Theta_1 - \ldots \\
&-I_{p-m+2} \cdot \Theta_m = I_{p+2} \\
&\qquad\qquad \vdots \\
&\alpha_p \cdot \Phi_n - I_{p+n-1} \cdot \Theta_1 - \ldots - I_{p+n-m} \cdot \Theta_m = I_{p+n}
\end{aligned} \quad (9)$$

3.1 Stage one estimation

In this stage initial parameter estimates are obtained based upon the inverse function operator (8).

Inverse function estimation

Consider the model:

$$I(B, \mathbf{i}) \cdot X_t = e_t^{ar}(\mathbf{i}) \quad (10)$$

that corresponds to the process representation implied by (8). In this model $I(B, \mathbf{i})$ represents a finite (truncated) s-th $(s > p+n)$ order approximation (permitted by way of assumption A2) of the inverse function operator, $\mathbf{i}$ the corresponding parameter vector, and $e_t^{ar}(\mathbf{i})$ the model's one-step-ahead prediction error at time t.

An interval estimate of the inverse function parameter vector is obtained through the expressions:

$$\hat{\mathbf{i}} = \left(\sum_{t=s+1}^{N} \boldsymbol{\psi}_t^{ar} (\boldsymbol{\psi}_t^{ar})^T \right)^{-1} \cdot \left(\sum_{t=s+1}^{N} \boldsymbol{\psi}_t^{ar} X_t \right) \quad (11)$$

$$\hat{Cov}[\hat{\mathbf{i}}] = (\hat{\sigma}_e^{ar})^2 \cdot \left[\frac{1}{N-s} \sum_{t=s+1}^{N} \boldsymbol{\psi}_t^{ar} (\boldsymbol{\psi}_t^{ar})^T \right]^{-1} \quad (12)$$

$$(\hat{\sigma}_e^{ar})^2 = \frac{1}{N-s} \sum_{t=s+1}^{N} [e_t^{ar}(\hat{\mathbf{i}})]^2 \quad (13)$$

with $\boldsymbol{\psi}_t^{ar} \triangleq [X_{t-1}\, X_{t-2} \ldots X_{t-s}]^T$.

Initial parameter estimation

Initial parameter estimation is based upon expressions (9) that relate the fractional power and AR/MA parameters with the model's inverse function. Given inverse function estimates, these lead to:

$$[\ \mathbf{A}\ \ \mathbf{J}\ ]\cdot\mathbf{p}=\begin{bmatrix}\mathbf{j}_1-\mathbf{a}_1\\ \mathbf{j}_2\end{bmatrix}+\mathbf{e}\Leftrightarrow$$

$$\Leftrightarrow\mathbf{X}\cdot\mathbf{p}=\mathbf{y}+\mathbf{e} \tag{14}$$

where:

$$\mathbf{A}_{[(p+n)\times n]}(i,j)\triangleq\begin{Bmatrix}0 & i<j\\ 1 & i=j\\ 0 & i\geq p+1+j\\ \alpha_{i-j} & otherwise\end{Bmatrix} \tag{15}$$

$$\mathbf{J}_{[(p+n)\times m]}(i,j)\triangleq\begin{Bmatrix}0 & i\prec j\\ -1 & i=j\\ -\hat{I}_{i-j} & otherwise\end{Bmatrix} \tag{16}$$

$$\mathbf{p}\triangleq\left[\phi_1\ \ldots\ \phi_n\vdots\theta_1\ \ldots\ \theta_m\right]^T=\left[\boldsymbol{\phi}^T\vdots\boldsymbol{\theta}^T\right]^T \tag{17}$$

$$\mathbf{j}_1\triangleq[\ \hat{I}_1\ \ \ldots\ \ \hat{I}_p\ ]^T,\quad \mathbf{j}_2\triangleq[\ \hat{I}_{p+1}\ \ \ldots\ \ \hat{I}_{p+n}\ ]^T \tag{18}$$

$$\mathbf{a}_1=[\ \alpha_1\ \ \ldots\ \ \alpha_p\ ]^T \tag{19}$$

with $\mathbf{e}$ representing an error vector. Expressions (14) define a special-form regression problem that is nonlinear in the fractional power d but linear in the AR/MA parameter vector $\mathbf{p}$. The optimization of the regression cost function:

$$J(d,\mathbf{p})\triangleq\frac{1}{2}\mathbf{e}_k^T\mathbf{Q}_{k\times k}\mathbf{e}_k \tag{20}$$

may be then accomplished through a pseudo-linear two-step procedure, according to which the fractional power is varied through an appropriate search scheme and conditional, upon it, AR/MA parameter estimates are obtained as:

$$\hat{\mathbf{p}}(\hat{d})=\left(\mathbf{X}_{k\times(n+m)}^T\mathbf{Q}_{k\times k}\mathbf{X}_{k\times(n+m)}\right)^{-1}\mathbf{X}_{k\times(n+m)}^T\mathbf{Q}_{k\times k}\mathbf{y}_{k\times 1} \tag{21}$$

In the above k $(n+m<k<n+p)$ refers to the number of scalar equations of (14) actually used in the regression, while $\mathbf{Q}_{k\times k}$ represents a proper weighting matrix. The procedure is terminated once the minimum of $J(d,\mathbf{p})$ is achieved. The innovations variance is then estimated as:

$$\hat{\sigma}_a^2=\hat{\sigma}_e^2(\hat{d},\hat{\mathbf{p}})=\frac{1}{N-q}\sum_{t=q+1}^{N}e_t^2(\hat{d},\hat{\mathbf{p}}) \tag{22}$$

with $q\triangleq p+n+m$.

3.2 Stage two estimation

This stage aims at refining the estimates of stage one. Let:

$$\hat{\bar{\mathbf{p}}}_{i-1}\triangleq\left[\hat{d}_{i-1}\ \hat{\boldsymbol{\phi}}_{i-1}^T\ \hat{\boldsymbol{\theta}}_{i-1}^T\ (\hat{\sigma}_a^2)_{i-1}\right]^T$$

denote the vector of $ARFIMA$ parameter estimates obtained at iteration $i-1$, and initially equal to those provided by stage one. At iteration i these estimates are updated as follows:

Fractional power and AR parameter estimation

Assuming small perturbations in the MA parameter estimates during successive iterations, that is $\Theta(B,\hat{\bar{\mathbf{p}}}_{i-1})\approx\Theta(B,\hat{\bar{\mathbf{p}}}_i)$ the $ARFIMA$ model (5) may, at iteration i, be approximately expressed as:

$$e_t(\hat{\bar{\mathbf{p}}}_i)\approx\Phi(B,\hat{\bar{\mathbf{p}}}_i)\cdot(1-B)^{d_i}\cdot\bar{X}_t^{i-1} \tag{23}$$

with:

$$\bar{X}_t^{i-1}\triangleq X_t/\Theta(B,\hat{\bar{\mathbf{p}}}_{i-1}) \tag{24}$$

The model (23) is of the Fractionally Integrated AutoRegressive $[FIAR(n,d)]$ form, and its parameters may be estimated via a procedure similar to that of Stage 1. In this case equation (14) is such that $\mathbf{X}=\mathbf{A}$, $\mathbf{p}=\boldsymbol{\phi}$, and the weighting matrix $\mathbf{Q}_{k\times k}$ in (20) is selected equal to the corresponding submatrix of the estimated inverse function covariance (12). Due to the form of (14) in this case, this leads to optimal, in the sense of the Gauss-Markov theorem [8], estimates.

MA parameter and innovations variance estimation

The MA parameters are then updated by solving the linear regression problem [obtained from (14)]:

$$\mathbf{J}\cdot\boldsymbol{\theta}=\mathbf{y}^M+\mathbf{e}^M \tag{25}$$

with $\mathbf{e}^M$ denoting the regression error vector, and:

$$y_i^M=\hat{I}_i-\sum_{k=0}^{i}\hat{\alpha}_k\hat{\phi}_{i-k}\qquad(i=1,2,\cdots p+n) \tag{26}$$

where $\alpha_0\equiv 1$, $\alpha_k\equiv 0$ $(k>p)$, $\phi_0\equiv 1$, $\phi_k\equiv 0$ $(k>n)$. The innovations variance is updated through (22).

4. Numerical experiment

Consider the $ARFIMA(1,d,2)$ process with parameters indicated in Table 1. This process is characterized by a sharp spectral valley, owing to the proximity

Parameter	Actual	Estimate ± std. deviation	
		PL	ML
d	0.30	0.313 ± 0.086	0.301 ± 0.067
ϕ_1	0.60	0.582 ± 0.073	0.587 ± 0.062
θ_1	-0.40	-0.396 ± 0.032	-0.380 ± 0.032
θ_2	0.99	0.916 ± 0.031	0.865 ± 0.097
σ_a^2	1.00	0.950 ± 0.086	1.124 ± 0.136

Table 1. Monte Carlo estimation results by the Pseudo-Linear (PL) and Maximum Likelihood (ML) methods ($N = 300$; 20 runs).

of its complex conjugate pair of zeros to the unit circle (magnitude of 0.995).

Monte Carlo estimation results by the Pseudo-Linear (PL) ($p = 20, s = 30$) and frequency-domain Maximum Likelihood (ML) [9] methods are, based upon 300-sample-long data records, summarized in Table 1. Despite the relatively short data record length and the significantly higher computational complexity of the ML method, the performance characteristics of the two methods appear similar. The ML method provides a slight improvement in the fractional power estimate, while, quite interestingly, the PL method achieves a noticeable improvement in the MA parameter estimates.

These observations are additionally confirmed from the estimated spectra, which are contrasted to the theoretical process spectrum in Figure 1. From these it is indeed evident that the PL method achieves a significantly reduced estimation scatter in the neighborhood of the spectral valley, while also providing a few estimates that are practically indistinguishable from the theoretical curve in the proximity of the spectral minimum. Similar behavior is observed with longer ($N = 1,000$) data records and processes characterized by sharp spectral peaks [10].

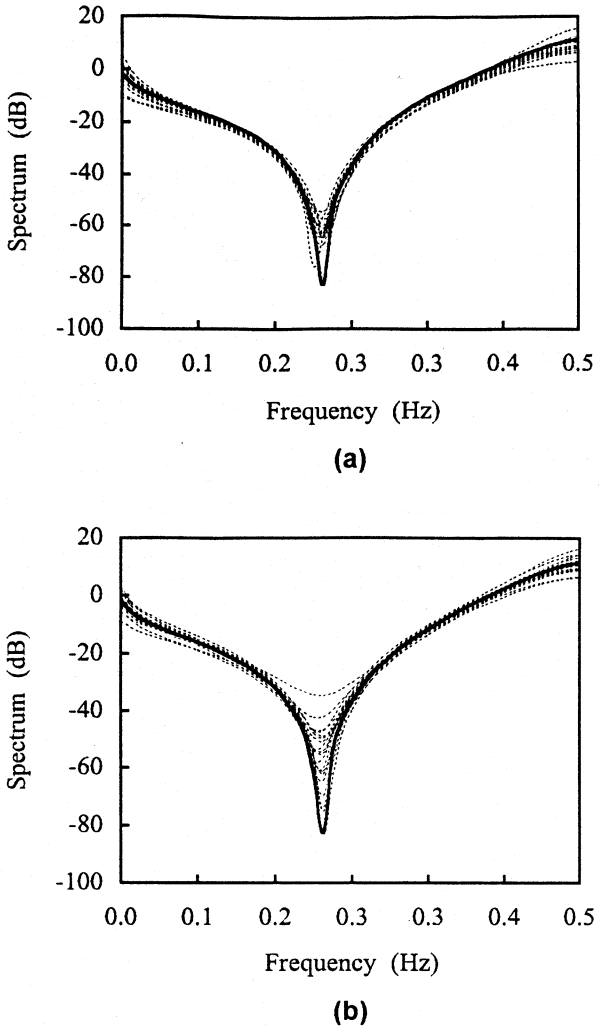

Figure 1. The actual (——) and estimated (- - - -) $ARFIMA(1, d, 2)$ spectra: (a) Pseudo-Linear method; (b) Maximum Likelihood method ($N = 300$; 20 runs).

References

[1] J.R.M. Hosking. Modeling persistence in hydrological time series using fractional differencing. *Water Resources Research*, 20:1898-1908, 1984.

[2] M. Deriche and A.H. Tewfik. Signal modeling with filtered discrete fractional noise processes. *IEEE Transactions on Signal Processing*, 41:2839-2849, 1993.

[3] R. Ben Mrad, S.D Fassois, J.A. Levitt and B.I. Bachrach. On-board prediction of power consumption in automotive active suspension systems - parts I and II. *Journal of Mechanical Systems and Signal Processing*, in press, 1996.

[4] G.E.P. Box, G.M. Jenkins and G.C. Reinsel. *Time Series Analysis: Forecasting and Control.* Third Edition, Prentice-Hall, 1994.

[5] R. Fox and M.S. Taqqu. Large sample properties for strongly dependent stationary gaussian time series. *Annals of Statistics*, 14:517-532, 1986.

[6] R. Dahlhaus. Small sample effects in time series analysis: a new asymptotic theory and a new estimate. *Annals of Statistics*, 16:808-841, 1988.

[7] F. Sowell. Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics*, 53:165-188, 1992.

[8] J.M. Mendel. *Lessons in Digital Estimation Theory.* Prentice-Hall, 1987.

[9] P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods.* Second Edition, Springer-Verlag, 1991.

[10] G.N. Fouskitakis. *Pseudo-Linear Estimation of Fractionally Integrated ARMA Processes with Applications.* Diploma Thesis, Department of Mechanical Engineering, University of Patras, 1995.