Full range autoregressive time series models

Dhanagopalan Venkatesan
Annamalai University, Tamil Nadu, India
sit_san@hotmail.com

Kolammal Sita Devi
Annamalai University, Tamil Nadu, India
sit_san@hotmail.com

Biagio Simonetti
Università del Sannio
simonetti@unisannio.it,

Abstract. *A new family of time series models, called the Full Range Autoregressive model, is introduced which avoids the difficult problem of order determination in time series analysis. Some of the basic statistical properties of the new model are studied.*

Keywords and Phrases. **Autoregressive model; autoregressive moving average model; order determination; FRAR model; identifiability; stationary condition.**

1. Introduction

The Autoregressive Moving Average (ARMA) models discussed in the literature are all of finite order type and so contains only a finite number of parameters. That is, they are generally based on the questionable assumption that the future value would be influenced only by a limited number of past values.
Further, most of the work in time series analysis are concerned with series having the property that the degree of dependence between observations, separated by a long time span, is zero or highly negligible. However, the empirical studies by Lawrance and Kottegoda (1977) reveal, particularly in cases arising in economics and hydrology, that the degree of dependence between observations a long time span apart, though small, is by no means negligible.
Long range dependence turns out to be widespread in nature and is characteristic of many hydrological, geophysical and economic records. Therefore, there is still a need for a family of models which can fully depict the properties of stationarity, linearity and long range dependence.
More over, the existing theory of autoregressive models assume that the coefficients of the model are not connected in any way among each other. That is, they are treated as unrelated constants. Therefore, it would be useful, from practical point of view, to propose new models which can accommodate long range dependence and have the property that the coefficients of the past values in the model are functions of a limited number of parameters. Thus, the chief objective of this paper is to introduce a family of new models which would involve a only few parameters and at the same time incorporate long range dependence which would be an acceptable alternative to the current models representing stationary time series.
A family of models, introduced in this paper, called Full Range Auto Regressive Model and denoted as FRAR model for short, are defined in such a way that they possess the following basic features.

1. The models should be capable of representing long term persistence. This is justified by the fact that the future may not depend on the present and a few past values alone, but may depend on the present and the whole past.

2. The models should be flexible enough to explain both the short-term and the long-term correlation structure of a series.
3. The parameters of the model, which are likely to be large in number due to (1), should exhibit some degree of dependence among themselves.

Therefore, the new models are expected to have infinite structure with a finite number of parameters and so completely avoid the problem of order determination.

An outline of the paper is as follows. In section 2 the FRAR model is defined, in section 3 the identifiability region for the FRAR model is obtained, in section 4 stationarity condition is derived and in section 5 the second order properties of the FRAR model is studied.

2. The Full Range Autoregressive Model

We define a family of models by a discrete-time stochastic process $\{x_t\}$, $t = 0, \pm 1, \pm 2, \pm 3, \ldots$, called the Full Range Auto Regressive (FRAR) model, by the difference equation

$$X_t = \sum_{r=1}^{\infty} a_r X_{t-r} + e_t , \qquad (1)$$

where $a_r = \dfrac{k \sin(r\theta) \cos(r\phi)}{\alpha^r}$, $r = 1, 2, 3, \ldots$

k, α, θ and ϕ are real parameters and e_1, e_2, e_3,….. are independent and identically distributed normal random variables with mean zero and variance σ^2. The initial assumptions about the parameters are as follows :

It is assumed that X_t will influence X_{t+n} for all positive n and the influence of X_t on X_{t+n} will decrease, at least for large n, and become insignificant as n becomes very large. Hence a_n must tend to zero as n goes to infinity. This is achieved by assuming that $\alpha > 1$. The feasibility of X_t having various magnitudes of influence on X_{t+n}, when n is small, is made possible by allowing k to take any real value. Because of the periodicity of the circular functions sine and cosine, the domain of θ and ϕ are restricted to the interval $(0, 2\pi)$.

Thus, the initial assumptions are $\alpha > 1$, $k \in R$, and θ, $\phi \in [0, 2\pi)$. i.e., $\Theta = (\alpha, k, \theta, \phi) \in S^*$, where $S^* = \{\alpha, k, \theta, \phi \,/\, \alpha > 1, k \in R, \ \theta, \phi \in [0, 2\pi)\}$. Further restrictions on the range of the parameters are placed by examining the identifiability of the model.

3. Identifiability Condition

Identifiability ensures that no two points in the parameter space could give the same stochastic model for the time series. In other words identifiability ensures that there is a one to one correspondence between the parameter space and set of associated probability models. Without identifiability it is meaningless to proceed to estimate the parameters of a model using a set of given data. In the present context, identifiability is achieved by restricting the parameters space in such a way that no two points in the parameter space could produce the same time series model.

The coefficients a_n's in (1) are functions of k, α, θ, ϕ as well as n. That is, $a_n = a_n (k, \alpha, \theta, \phi) = \left(\dfrac{k}{\alpha^n}\right) \sin (n\theta) \cos(n\phi)$, $\theta \in S^*$, $n = 1, 2, \ldots$.

Define
$A = \{\alpha, k, \theta, \phi \ / \alpha > 1, k \in R, \pi \le \theta, \phi < 2\pi\}$,
$B = \{\alpha, k, \theta, \phi \ / \alpha > 1, k \in R, 0 \le \theta < \pi, \pi \le \phi < 2\pi\}$,
$C = \{\alpha, k, \theta, \phi \ / \alpha > 1, k \in R, \pi \le \theta < 2\pi, 0 \le \phi < \pi\}$,
$D = \{\alpha, k, \theta, \phi \ / \alpha > 1, k \in R, 0 \le \theta, \phi < \pi\}$.

Since $a_n (k, \alpha, \theta, \phi) = a_n (-k, \alpha, 2\pi - \theta, 2\pi - \phi)$, $\theta \in S^*$ (2)

To each $(\alpha, k, \theta, \phi)$ belonging to A there is a $(\alpha, k, \theta', \phi')$ $(\theta' = 2\pi - \theta$ and $\phi' = 2\pi - \phi)$ belonging to D such that $a_n(k, \alpha, \theta, \phi) = a_n(-k, \alpha, \theta', \phi')$. So A is omitted. Similarly, it can be shown that B and C can also be omitted.

Define $D_1 = \{\alpha, k, \theta, \phi \;/\; \alpha > 1, k \in R, \pi/2 \le \theta, \phi < \pi\}$,

$$D_2 = \{\alpha, k, \theta, \phi \;/\; \alpha > 1, k \in R, 0 \le \theta < \frac{\pi}{2}, \frac{\pi}{2} \le \phi < \pi\},$$

$$D_3 = \{\alpha, k, \theta, \phi \;/\; \alpha > 1, k \in R, 0 \le \theta, \phi < \pi/2\},$$

$$D_4 = \{\alpha, k, \theta, \phi \;/\; \alpha > 1, k \in R, \frac{\pi}{2} \le \theta < \pi,\; 0 \le \phi < \frac{\pi}{2}\}.$$

Since $a_n(k, \alpha, \theta, \phi) = a_n(-k, \alpha, \pi - \theta, \pi - \phi)$ for $k \in R, \alpha > 1, 0 \le \theta, \phi \le \pi$ (3)

Using (3) it can be shown as before, that the regions D_1 and D_2 can be omitted. Since no further reduction is possible, it is finally deduced that the region of identifiability of the model is given by $S = \{\alpha, k, \theta, \phi \;/\; k \in R, \alpha > 1, \theta \in [0,\pi)$ and $\phi \in [0, \pi/2)\}$.

4. Stationarity of the FRAR process

The stationarity of the newly developed FRAR time series model is now examined. The model is given by

$X_t = \sum a_r X_{t-r} + e_t$

That is, $(1 - a_1 B - a_2 B^2 - \ldots) X_t = e_t$,

Where B is the backward shift operator, defined by $B^n X_t = X_{t-n}$

Here and in the following pages $\sum$ stands for summation with respect to r from 1 through infinity unless otherwise stated.

Thus the model is given by

$\psi(B) X_t = e_t$, or $X_t = \psi^{-1}(B) e_t$

Where $\psi(B) = 1 - a_1B - a_2 B^2 - a_3 B^3 - \ldots\ldots$

Box and Jenkins (1976) and Priestley (1981) have shown that a necessary condition for the stationarity of such processes is that the roots of the equation $\psi(B) = 0$ must all lie outside the unit circle. So, it is now proposed to investigate the nature of the zeros of $\psi(B)$.

The power series $\psi(B)$ may be rewritten as

$\psi(B) = 1 - [a_1B + a_2B^2 + a_3B^3 + \ldots] = 1 - \sum a_nB^n$,

where $a_nB^n = (k B^n / \alpha^n) \sin(n\theta) \cos(n\phi) = (k' B^n / \alpha^n)(\sin n\theta_1 + \sin n\theta_2)$

where $k' = k/2$, $\theta_1 = \theta + \phi$ and $\theta_2 = \theta - \phi$.

Therefore,

$$\sum a_n B^n = \sum \frac{k' B^n}{\alpha^n} \sin(n\theta_1) + \sum \frac{k' B^n}{\alpha^n} \sin(n\theta_2) .$$

The above two series are separately evaluated below.

$$\sum \frac{k' B^n}{n} \sin(n\theta_1) = \text{IP of } \sum \frac{k' B^n}{\alpha^n} e^{in\theta_1} = \text{IP}\{KB e^{i\theta_1} (\alpha - Be^{i\theta_1})^{-1}\}$$

$= (k' B \alpha \sin\theta_1) / G_1$, $G_1 = B^2 + \alpha^2 - 2 B \alpha \cos\theta_1$

Where IP stands for imaginary part. Similarly, it can be shown that

$$\sum \frac{k' B^n}{\alpha^n} \sin(n\theta_2) = (k' B \alpha \sin\theta_2) / G_2$$

where $G_2 = B^2 + \alpha^2 - 2B \alpha \cos\theta_2$.

Therefore, $\sum a_nB^n$

$= k'B\alpha [(B^2 + \alpha^2)(\sin\theta_1 + \sin\theta_2) - 2B\alpha(\sin\theta_1 \cos\theta_2 - \cos\theta_1 \sin\theta_2)] / G_1G_2$

Thus, $\psi(B) = 1 - \Sigma a_nB^n = 0$ implies that

$[(B^2 + \alpha^2 - 2B\alpha \cos\theta_1) (B^2 + \alpha^2 - 2B\alpha \cos\theta_2)]$
$- k' B\alpha [(B^2 + \alpha^2) (\sin \theta_1 + \sin \theta_2) - 2B\alpha \sin 2\phi] = 0.$
That is,
$[(B^2 + \alpha^2 - 2B\alpha \cos\theta_1) (B^2 + \alpha^2 - 2B\alpha \cos\theta_2)] - k' B\alpha [(B^2 + \alpha^2) s_1 - 2Bd_1c_2] = 0$
where
$s_1 = \sin\theta_1 + \sin\theta_2 = 2 \sin\theta \cos\phi$; $c_2 = \sin 2\theta_2$,
$c_1 = \cos\theta_1 + \cos\theta_2 = 2 \cos\theta \cos\phi$; $d_1 = \cos\theta_1 - \cos\theta_2$.
After simplifying, the above equation becomes
$B^4 - B^3 \alpha (2c_1 + k' s_1) + B^2\alpha^2 (2 + 4 d_1 + 2 k' c_2) - B\alpha^3 (2c_1 + k' s_1) + \alpha^4 = 0.$
That is,
$$B^4 - B^3 \alpha A_1 + B^2 \alpha^2 A_2 - B \alpha^3 A_1 + \alpha^4 = 0 \qquad (4)$$
$$\text{or} \quad S^4 - A_1 S^3 + A_2 S^2 - A_1 S + 1 = 0 \qquad (5)$$
where $S = B / \alpha$,
$A_1 = 2c_1 + k' s_1 = \cos\phi (4 \cos\theta + k \sin\theta)$, and
$A_2 = 2 + 4d_1 + 2k' c_2 = 2 [1 - \sin\phi (4 \sin\theta - k \cos \phi)]$,
This is a reciprocal equation of degree 4 which reduces to
$Z^2 - A_1 Z + (A_2 - 2) = 0$ where $Z = S + (1 / S)$.
The roots of this equation are, say r_1 and r_2, are given by
$$Z = (1/2) \left[A_1 \pm \sqrt{(A_1^2 - 4A_2 + 8}\right]$$
Since $Z = S + (1/ S)$, one finally gets the four roots of the equation (4), as
$$R_1 = (1/2) \left[r_1 + \sqrt{r_1^2 - 4}\right] \quad , \qquad R_2 = (1/2) \left[r_1 - \sqrt{r_1^2 - 4}\right]$$
$$R_3 = (1/2) \left[r_2 + \sqrt{r_2^2 - 4}\right] \qquad \text{and} \quad R_4 = (1/2) \left[r_2 - \sqrt{r_2^2 - 4}\right]$$
All the roots are complicated functions of $\sin\theta$, $\cos\theta$, $\sin\phi$ and $\cos\phi$. Therefore it is very difficult to study the nature of the roots. However, since equation (5) is a reciprocal equation, if S_0 is a root of the equation (5) then $(1/S_0)$ is also a root. This implies that αS_0 and (α / S_0) are roots of equation (4). But α can take any value bigger than 1. So, for sufficient large α the process is stationary. But when α is small it seems difficult to examine the stationarity of the process by this approach. Hence, it is proposed to study the second order properties of the process in the following section.

5. Second order properties of the FRAR Process

To examine the second order properties of the new model proposed one has to solve the different equation (1), so as to obtain an expression for X_t in terms of $e_t, e_{t-1}, e_{t-2}, e_{t-3}, \ldots$.

The precise solution of this equation depends on the initial conditions. So to investigate the nature of the first and second moments of X_t, following Priestley (1981), it is assumed that $X_t = 0$ for $t < -N$, N being the number of observations in the time series. Then solving (1) by repeated substitutions one obtains

$X_t = e_t + a_{11} X_{t-1} + a_{12} X_{t-2} + a_{13} X_{t-3} + \ldots$

Where $a_{1j} = a_j = (k / \alpha^j) \sin (j\theta) \cos(j\phi)$; $j = 1, 2, \ldots$

$= e_t + a_{11} e_{t-1} + a_{22} X_{t-2} + a_{23} X_{t-3} + a_{24} X_{t-4} + \ldots$

where $a_{2j} = a_{11} a_1 j\text{-}1 + a_{1j}$, $j = 2, 3, 4 \ldots$

Similarly proceeding one finally gets

$X_t = [e_t + a_{11} e_{t-1} + a_{22} e_{t-2} + a_{33} e_{t-3} + a_{44} e_{t-4} + \ldots + a_{pp} e_{t-p}]$
$+ [a_{p+1\ p+1} X_{t-(p+1)} + a_{p+1\ p+2} X_{t-(p+2)} + \ldots\ldots]$

where $a_{ij} = a_{i-1\ i-1}\ a_{1\ j+1-i} + a_{i-1\ j}$, $j > i = 2, 3, \ldots$

Thus, if it is assumed that $X_t = 0$ for $t \le -N$, which implies has $n = N + t - 1$, then, $X_t = e_t + a_{11} e_{t-1} + a_{22} e_{t-2} + a_{33} e_{t-3} + a_{44} e_{t-4} + \ldots + a_{N+t-1,\ N+t-1}\, e_{1-N}$

Further, it can be shown that

$$E[X_t X_{t+1}] = \sigma_e^2 \left[a_{11}\left(1 + a_{11}^2 + a_{22}^2 + \ldots + a_{N+t-1\ N+t-1}^2\right) + \left(a_{11} a_{12} + a_{22} a_{23} + \ldots\ldots + a_{N+t-2\ N+t-2}\ a_{N+t-2\ N+t-1}\right)\right]$$

Similarly

$$E[X_t X_{t+2}] = \sigma_e^2 \left[a_{22}\left(1 + a_{11}^2 + a_{22}^2 + \ldots + a_{N+t-1\ N+t-1}^2\right) + a_{11}\left(a_{11} a_{12} + a_{22} a_{23} + \ldots + a_{N+t-2\ N+t-2}\ a_{N+t-2\ N+t-1}\right)\right] + \left(a_{11} a_{13} + a_{22} a_{24} + \ldots + a_{N+t-3\ N+t-3}\ a_{N+t-3\ N+t-1}\right)\Big]$$

Similarly it can be shown that

$$E[X_t X_{t+3}] = \sigma_e^2 \left[a_{33}\left(1 + a_{11}^2\ (1 + a_{22} + a_{44} + \ldots + a_{N+t-3\ N+t-3})\right) + a_{11}\left(a_{22} a_{34} + a_{33} a_{45} + \ldots + a_{N+t-3\ N+t-3}\ a_{N+t\ N+t-1}\right)\right] + \left(a_{11} a_{34} + a_{22} a_{45} + \ldots + a_{N+t-1\ N+t-1}\ a_{N+t-3\ N+t-3}\right)\Big]$$

and in general

$$E[X_t X_{t+s}] = \sigma_e^2\left[a_{ss} + a_{11} a_{s+1\ s+1} + a_{22} a_{s+2\ s+2} + \ldots\ldots + a_{N+t-1\ N+t-1} a_{N+t+s-1\ N+t+s-1}\right]$$

Where $a_{ss} = a_{11}\ a_{s-1\ s-1} + a_{s-1\ s}$.

Therefore allowing $N \to \infty$, we get

$$E[X_t] = 0,$$

$$Var[X_t] = \sigma_e^2\ \left[1 + a_{11}^2 + a_{22}^2 + \ldots\ldots.\right] \text{ and}$$

$$E[X_t\ X_{t+s}] = \sigma_e^2\ \left[a_{ss} + a_{11}\ a_{s+1\ s+1 + \ldots}\right]$$

provided the series on the right converges. Thus, it is seen that if $E(X_t X_{t+s})$ exists then it is a function of s only.

In order to examine the asymptotic behavior of the second order moments of X_t, first the behavior of a_{ij}, as j tends infinity, is investigated. Since $a_{1j} = a_j = (k / \alpha^j) \sin (j\theta) \cos(j\phi)$, $|a_{1j}| \le \frac{|k|}{\alpha^j}$. Similarly,

$$|a_{2j}| \le \frac{|k|}{\alpha^j} + \frac{|k|^2}{\alpha^j} \le \frac{|k|\,(1+|k|)}{a^j};\ j \ge 2.$$

Thus, in general

$$|a_{nj}| \le \frac{|k|\,(1+|k|)^{n-1}}{\alpha^j};\quad \text{for } j \ge n.$$

Since $\alpha > 1$, the above relation implies that $|a_{nj}| \to 0$ as $j \to \infty$, for any fixed n.

Further $|a_{jj}| \le \frac{|k|\,(1+|k|)^{j-1}}{\alpha^j}$.

Thus $\sum a_{jj}^2$ will converge if $\left|\frac{(1+|k|)}{\alpha}\right| < 1$,

that is, if $(1 + |k|) < \alpha$ or $|k| < \alpha - 1$ or $1 - \alpha < k < \alpha - 1$.

The auto-correlation function of the FRAR process is obtained as

$$\rho(s) = \frac{R(s)}{R(0)} = \frac{E(x_t\ x_{t+s})}{E(X^2)} = \frac{\left[a_{ss} + a_{11}\ a_{s+1\ s+1} + a_{22}\ a_{s+2\ s+2} + \ldots + a_{N+t-1\ N+t-1}\ a_{N+t+s-1\ \ N+t+s-1}\right]}{\left[1 + a_{11}^2 + a_{22}^2 + \ldots + a_{N+t-1\ N+t-1}^2\right]}$$

If we make use of $|a_{jj}| \le \frac{|k|\,(1+|k|)^{j-1}}{\alpha^j}$.

then Var (X_t) = $\sigma^2_{x_t} = \sigma^2_e \left[\sum_{j=0}^{\infty} a^2_{jj}\right]$ with $a_{00} = 1$

$$\leq \sigma^2_e \frac{k^2}{(1+k)^2}\left[\frac{\alpha^2}{\alpha^2-(1+k)^2}\right]$$

Thus, when $\frac{1+|k|}{\alpha} \leq 1$, $\sigma^2_{x_t}$ exist and is finite.

Similarly,

$$E[X_t\, X_{t+s}] = \sigma^2_e\left[a_{ss} + a_{11}\, a_{s+1\, s+1} + a_{22}\, a_{s+2\, s+2} + ...\right] \leq \sigma^2_e\, \frac{k(1+k)^{s-1}}{\alpha^s} \sum_{j=0}^{\infty} a^2_{jj} \quad ; \quad a_{00} = 1.$$

So $\left|E[X_t\, X_{t+s}]\right| \leq \sigma^2_e \frac{k^2}{(1+k)^2} \cdot \frac{k(1+k)^{s-1}}{\alpha^s}\left[\frac{\alpha^2}{\alpha^2-(1+k)^2}\right].$

Therefore, the auto-correlation function of the process exists and, as shown earlier, it is a function of s only. Finally allowing $t \to \infty$, it is seen that

(i) $\lim_{t\to\infty} E(X_t)$ and $\lim_{t\to\infty} Var(X_t)$ exist finitely.

(ii) $\lim_{t\to\infty}$ Cov (X_t, X_{t+s}) exists finitely and is a function of ‘s’ only.

Therefore, it is seen that the newly proposed FRAR model is asymptotically stationary at least up to order 2 provided $1 - \alpha < k < \alpha - 1$.

Thus, the new FRAR model incorporates long range dependence, involves only four parameters and is totally free from order determination problems.

References

Box, G.E.P. and Jenkins, J.M., (1976). Time Series Analysis, Forecasting, and Control, Holden Day, San Francisco.

Lawrance, A.J. and N.T. Kottegoda., (1977). “Stochastic Modelling river flow time series (with discussion)”, Journal of the Royal Statistical Society, Series A, 140, 1 – 47.

Priestley, M.B., (1981). Spectral Analysis and Time Series, Academic Press, London and New York.