Indian Journal of Science and Technology Vol. 3 No. 9 (Sep 2010) ISSN: 0974- 6846

Application of time series models

P. Sekar

Department of Mathematics, Pachaiyappa's College for Men, Kancheepuram-631 501, TN, India
cicesekar@yahoo.co.in; icesekar@rediffmail.com

Abstract

Forecasting is an ultimate aim in the study of time series analysis. Anyone who is engaged in planning, controlling and managing projects, personnel, finance and operations will be interested in knowing what will happen in future with the analysis of the available data.

Keywords: Time series, ARMA, ARIMA, ARARMA, fractional differencing.

General

Some of the studies made in the time series analysis are illustrated with the following applications of time series data. i). ARIMA model (Box & Jenkins, 1976) is fit into the monthly data of residential electricity usage in IOWA city (1971-1979). ii). ARARMA model (Parzen, 1982) is fit into the UK air-line data for the period from January 1964 to December 1970 (Montgomary & Johnson, 1976). iii). The monthly mean rainfall data of Coimbatore city are taken from the month of January 1984 to the month of December 1994 to apply fractional ARIMA modeling (Hosking, 1981). iv). Models are identified simultaneously for the Chennai city traffic accidents data from the month of January 1987 to December 1997.

Arima modeling

Analysis of monthly residential electricity usage in IOWA city

The monthly average residential electricity usage in IOWA city from the period (1971-1979) is taken for analysis.

The logged series is shown in Table 1. Autocorrelation of the following series are estimated.

i. The original logged series, z_t;
ii. The logged series differenced with respect to months only, ∇z_t
iii. The logged series differenced with respect to years only, $\nabla_{12} z_t$
iv. The logged series differenced with respect to months and years, $\nabla\nabla_{12} z_t$.
v. The autocorrelation of z_t, ∇z_t, $\nabla_{12} z_t$, $\nabla\nabla_{12} z_t$ are shown in Table 2.

The autocorrelations for z_t's are large and fail to die out at higher lags. While simple differencing reduces the correlation in general, a very heavy periodic component remains. This is evidenced particularly by very large autocorrelations at lags 12, 24, 36 and 48. Simple differencing with respect to period 12 results in autocorrelations which are first persistently positive and persistently negative. The differencing $\nabla\nabla_{12}$ markedly reduces autocorrelations throughout. The large autocorrelations at lag 1 & lag 12 indicate that the model will have one regular moving average operator and seasonal moving average operator using the hypothesis that autocorrelation function of a moving average process of order q has a cut off after lag q, while its partial autocorrelation fails off.

The model can be written in the form

$$w_t = (1-\theta B)(1-\Theta B^{12})\, a_t,\quad w_t = \nabla\nabla_{12} z_t$$
$$= (1 - \theta B - \Theta B^{12} + \theta\Theta B^{13})\, a_t$$
$$w_{t+k} = a_{t+k} - \theta\, a_{t+k-1} - \Theta\, a_{t+k-12} + \theta\Theta a_{t+k-13}$$

The autocovariance at lag k is defined as

$$\gamma_k = E[w_t w_{t+k}]$$
$$\gamma_0 = E[w_t^2]$$
$$= (1 + \theta^2 + \Theta^2 + \theta^2\Theta^2)\,\sigma_a^2 \text{ where } \sigma_a^2 = E(a_t^2)$$
$$= (1+\theta^2)(1+\Theta^2)\,\sigma_a^2$$
$$\gamma_1 = (-\theta-\theta\Theta^2)\,\sigma_a^2 = -\theta(1+\Theta^2)\,\sigma_a^2$$
$$\gamma_{11} = \theta\Theta\sigma_a^2$$
$$\gamma_{12} = [-\Theta-\Theta\theta^2]\,\sigma_a^2 = -\Theta(1+\theta^2)\,\sigma_a^2$$
$$\gamma_{13} = \theta\Theta\sigma_a^2$$

The autocorrelation at lag k is defined as

$$\rho_k = \frac{\gamma_k}{\gamma_0}$$

$$\rho_1 = \frac{-\theta(1+\Theta^2)\sigma_a^2}{(1+\theta^2)((1+\Theta^2)\sigma_a^2} = \frac{-\theta}{1+\theta^2} \qquad (1.1)$$

$$\rho_{12} = \frac{\gamma_{12}}{\gamma_0} = \frac{-\Theta(1+\theta^2)\sigma_a^2}{(1+\theta^2)((1+\Theta^2)\sigma_a^2} = \frac{-\Theta}{1+\Theta^2} \qquad (1.2)$$

It is found that for the airline data

$\rho_1 = -0.40$, $\rho_{12} = -0.339$

Substitute the values of ρ_1 and ρ_{12} in 1.1 and 1.2

We obtain
$$0.40\,\theta^2 - \theta + 0.40 = 0$$
$$0.339\,\Theta^2 - \Theta + 0.339 = 0$$

Solving these equations we obtain the initial estimates for θ and Θ

$$\hat{\theta} = 0.50,\ \hat{\Theta} = 0.3907$$

Indian Journal of Science and Technology Vol. 3 No. 9 (Sep 2010) ISSN: 0974- 6846

Table 1. The logged data of monthly average electricity usage in IOWA city (1971-79).

t	z_t	t	z_t	t	z_t
1	6.1180970	37	6.1800170	73	6.2728770
2	6.0426330	38	6.0661080	74	6.2285110
3	5.9635790	39	5.9964520	75	0.0776420
4	5.9080830	40	5.9889610	76	6.0088130
5	6.1312270	41	5.9242560	77	5.9712620
6	5.9558370	42	6.0637850	78	6.2747620
7	6.2166060	43	6.5012900	79	6.5652650
8	6.4068800	44	6.4967750	80	6.4892050
9	6.2897160	45	6.2166060	81	6.2146080
10	5.9428000	46	5.9712620	82	6.0258660
11	5.9814140	47	5.9914650	83	6.0354810
12	6.1136820	48	6.2166060	84	6.2538290
13	6.1527330	49	6.2971090	85	6.2822670
14	6.1841490	50	6.1590950	86	6.2205900
15	6.1224930	51	6.1114670	87	6.1398840
16	6.0258660	52	6.0958250	88	6.0258660
17	5.9610050	53	5.9506430	89	5.9480350
18	6.2186000	54	6.1923620	90	6.1569790
19	6.2971090	55	6.6372580	91	6.5161930
20	6.4645880	56	6.5666720	92	6.4329400
21	6.4248690	57	6.3935910	93	6.4800440
22	6.1180970	58	6.0063530	94	6.1612070
23	6.0614570	59	6.0282780	95	6.0450050
24	6.1590950	60	6.0867750	96	6.2166060
25	6.2225760	61	6.2595820	97	6.2841340
26	6.1180970	62	6.2186000	98	6.2265370
27	6.0497340	63	6.0844990	99	6.1675170
28	6.0063530	64	6.0402550	100	6.0450050
29	5.9506430	65	5.9584250	101	5.9532430
30	6.1441860	66	6.1158920	102	6.0822190
31	6.5042880	67	6.4457200	103	6.3026190
32	6.4952650	68	6.4567690	104	6.4645880
33	6.5102580	69	6.3561080	105	6.3385940
34	6.0544400	70	6.0185930	106	5.9889610
35	6.0282780	71	6.1202970		
36	6.0913100	72	6.2383250		

Estimation of parameters

Corresponding to N = nd - sD, where d is the order of regular differencing, D is the order of seasonal differencing with period `s' the unconditional log - likelihood for the multiplicative model (p,d,q) x $(P,D,Q)_{12}$ is given by

$$l(\beta, \sigma_a) = f(\beta) - n \log \sigma_a - \frac{s(\beta)}{2\sigma_a^2}$$

Where β is a general symbol for k = p+q+P+Q parameters.

The unconditional sum of squares function is given by:

$s(\beta) = \sum_{t=-\infty}^{\infty}$ $[a_t / \beta, w]$ denotes the expectation of a_t conditional on β and w.

Table 2. Autocorrelation for various differences.

lagk	ACF of z_t	ACF of ∇z_t	ACF of $\nabla_{12} z_t$	ACF of $\nabla\nabla_{12} z_t$
1	0.517	0.225	0.329	-0.400
2	-0.163	-0.353	0.183	0.001
3	-0.500	-0.541	0.068	-0.082
4	-0.331	-0.240	0.012	0.036
5	0.060	0.175	-0.056	-0.156
6	0.273	0.412	0.083	0.074
7	0.102	0.261	0.133	0.115
8	-0.308	-0.231	0.035	-0.010
9	-0.496	-0.528	-0.056	0.018
10	-0.183	-0.315	-0.173	-0.069
11	0.428	0.276	-0.206	0.124
12	0.768	0.709	-0.390	-0.339
13	0.438	0.305	-0.109	0.251
14	-0.165	-0.320	-0.125	-0.232
15	-0.466	-0.476	0.100	0.256
16	-0.309	-0.217	-0.010	-0.097
17	0.035	0.177	-0.006	0.084
18	0.203	0.325	-0.093	-0.090
19	0.066	0.231	-0.061	0.020
20	-0.282	-0.189	-0.043	-0.101
21	-0.446	-0.499	0.091	0.048
22	-0.136	-0.243	0.170	0.137
23	0.404	0.257	0.070	-0.012
24	0.688	0.629	-0.033	-0.061
25	0.373	0.227	-0.040	0.016

For moderate large value of n, $l(\beta,\sigma_a)$ is dominated by

$$\frac{s(\beta)}{2\sigma_a^2}$$

The contours of the unconditional sum of square function in the space parameters β are very nearly contours of the log likelihood.

Calculation of the unconditional sum of squares s(β) for the multiplicative model (0,1,1) x $(0,1,1)_{12}$

The multiplicative model (0, 1, 1) x (0, 1, 1) $_{12}$ may be written in the form

w_t = $(1-\Theta B)(1-\Theta B^{12})\, a_t$ or

w_t = $(1-\Theta F)(1-\Theta F^{12})\, a_t$

Where F is the forward shift operator.

Assume that

$E(w_t) = \mu = o$.

We can write

$[e_t] = [w_t] + \theta [e_{t+1}] + \Theta [e_{t+12}] - \theta \Theta [e_{t+13}]$

$[a_t] = [w_t] + \theta [a_{t-1}] + \Theta [a_{t-12}] - \theta \Theta [a_{t-13}]$

where $[w_t] = w_t$ for t = 1, 2,.... n.

i. Assume that $[e_t] = 0$ if t >n

ii. Knowing the values of w_t, the values of $[e_t]$'s are calculated recursively up to $[e_t]$

iii. The values of $[e_0]$, $[e_{-1}]$, ..., $[e_{-12}]$ are distributed independently of w_1, w_2,....

iv. Knowing all the values of e_{-12}, e_{-11}.... e_0 .... e_{93}

We back forecast the values of w_{-12}, w_{-11},... w_0 (w_{-j} > 0 for j > 12).

v. The values of $[a_t]$ for t = -12, -11, ... 0,1,2,.... 13 recursively ($a_{-j} > 0$ for j > 12) are obtained.

vi. The sum of squares

$$s(\beta) = \sum_{t=-\infty}^{\infty} [a_t / \beta, w]^2 \text{ for different values of}$$

parameter β = (θ, Θ) are obtained.

vii. The parameter estimation obtained by minimizing the sum of squares s(β) give least square estimates (Table 3).

The least square estimates for the seasonal multiplicative model (0, 1, 1) x $(0, 1, 1)_{12}$ of monthly average residential electricity usage in IOWA city (1971-1979) data are

$\hat{\theta} = 0.77$, $\hat{\Theta} = 0.84$.

Table 3. Least square estimates for the multiplicative model.

θ	Θ	Sum of squares s(β)
0.750	0.870	0.6119
0.750	0.880	0.6156
0.750	0.890	0.6212
0.760	0.800	0.6127
0.760	0.810	0.6110
0.760	0.820	0.6097
0.760	0.830	0.6088
0.760	0.840	0.6084
0.760	0.850	0.6087
0.760	0.860	0.6099
0.760	0.870	0.6124
0.760	0.880	0.6164
0.760	0.890	0.6225
0.770	0.800	0.6122
0.770	0.810	0.6106
0.770	0.820	0.6094
0.770	0.830	0.6086
0.770	0.840	0.6083
0.770	0.850	0.6088
0.770	0.860	0.6104
0.770	0.870	0.6131
0.770	0.880	0.6176
0.770	0.890	0.6242

1.2.4. Iterative calculation of least square estimates

For the multiplicative model (0, 1, 1), x $(0, 1, 1)_{12}$, we have

$a_{t,0} = (\theta - \theta_0)\, x_{1,t} + (\Theta - \Theta_o)\, x_{2,t} + a_t$

where

$$x_{1,t} = -\frac{\partial a_t}{\partial \theta}, \quad x_{2,t} = \frac{\partial a_t}{\partial \Theta}$$

and θ, Θ are assigned initial values. The derivatives $x_{1,t}$, $x_{2,t}$ are calculated numerically.

For the average residential electricity usage for the data, taking the preliminary estimates $\hat{\theta} = 0.50$, $\hat{\Theta} = 0.3907$ and using the above process the iterative estimation of θ and Θ are tabulated as in Table 4.

Table 4. Iterative estimation of θ & Θ.

Iteration	θ	Θ
1	0.6468	0.7209
2	0.6817	0.8450
3	0.6897	0.8922
4	0.6915	0.8836
5	0.6919	0.8880
6	0.6920	0.8859
7	0.6920	0.8869
8	0.6920	0.8865
9	0.6920	0.8867
10	0.6920	0.8866
11	0.6920	0.8866
12	0.6920	0.8866
13	0.6920	0.8866
14	0.6920	0.8866
15	0.6920	0.8866

The least square estimates are

$\hat{\theta} = 0.6920$

$\hat{\Theta} = 0.8866$

Cumulative periodogram check

If the model were adequate and parameter known exactly, the plot of $C(f_j)$ against f_j would be scattered about a straight line joining the points (0,0) and (0.5,1). For the model identified using ARIMA multiplicative model $C(f_j)$'s are found and the plot of $C(f_j)$ against f_j will be useful.

The Kolmogorov-Smirnov 5% and 25% probability limits supplying a very rough guide to the significance of apparent deviation fail in this instance to indicate any significant departure from the assumed model. The limit lines are drawn at distances $\pm K_\varepsilon / \sqrt{q}$.

The cumulative periodogram are drawn for the models having least square estimates $\hat{\theta} = 0.692$, $\hat{\Theta} = 0.8866$ and $\hat{\theta} = 0.77$, $\hat{\Theta} = 0.84$. The 5% limit lines inserted deviate from the theoretical lines by $\pm 1.36 / \sqrt{51} = \pm 0.1904$. The 25% limit lines drawn deviate from the theoretical lines by $\pm 1.02/\sqrt{51} = \pm 0.1428$. Therefore the least square estimates are taken to be $\hat{\theta} = 0.692$, $\hat{\Theta} = 0.8866$.

Forecasting

After identifying the model and estimating the parameter, the diagnostic checking is done. If the model is adequate, then the model is used for forecasting. Forecasts are computed from the difference equation itself. The value of z_{t+l} is forecast for $l \geq 1$, when it is currently standing at time t.

The multiplicative seasonal model (Box & Jenkins 1976) for the air-line data is written in the form

$$z_{t-l} = z_{t+l-1} + z_{t+l-12} - z_{t+l-13} + a_{t+l} - \theta a_{t+l-1} - \Theta\, a_{t+l-12} + \theta\Theta a_{t+l-13}$$

where θ = 0.692, Θ = 0.8866

The minimum mean square error forecast at lead time l at origin t is given by

$$\hat{z}_t(l) = [z_{t+l-1} + z_{t+l-12} - z_{t+l-13} + a_{t+l} - \theta\, a_{t+l-1} - \Theta\, a_{t+l-12} + \theta\Theta\, a_{t+l-13}]$$

Also (i). $[z_{t+l}] = E(z_{t+l} / \theta, \Theta, zt, z_{t-1} ....)$ is the conditional expectation of z_{t+l} taken at origin t.

(ii).Invertible model fitted to actual data usually yield forecast which depend only on recent value of the series.

(iii). The forecast are insensitive to small changes in parameter values such as the one introduced by estimation errors.

(iv). $$[z_{t+j}] = \begin{cases} z_{t+j}, & j \leq 0 \\ \hat{z}_t(j), & j > 0 \end{cases}$$

$$[a_{t+j}] = \begin{cases} a_{t+j}, & j \leq 0 \\ 0, & j > 0 \end{cases}$$

The unknown z_t's are replaced by forecasts and the unknown a_t's are replaced by zeroes.

Indian Journal of Science and Technology Vol. 3 No. 9 (Sep 2010) ISSN: 0974- 6846

ARARMA modeling

1.3.1. Analysis of airline data of total miles flown in the UK for the period 1964-1970

Table 5. The logged airline data of united kingdom (1964-1970).

t	Zt	t	Zt	t	Zt
1	1.9836000	29	2.3275000	57	2.6255000
2	1.9132000	30	2.5081000	58	2.3848000
3	2.0566000	31	2.4542000	59	2.2289000
4	2.1248000	32	2.4490000	60	2.3911000
5	2.2049000	33	2.5191000	61	2.2503000
6	2.3271000	34	2.2340000	62	2.1882000
7	2.4006000	35	2.0911000	63	2.4516000
8	2.3871000	36	2.2279000	64	2.1806000
9	2.3353000	37	2.1203000	65	2.5287000
10	2.2093000	38	2.0667000	66	2.6919000
11	2.0393000	39	2.3020000	67	2.6150000
12	2.0284000	40	2.3104000	68	2.6197000
13	2.1223000	41	2.3796000	69	2.7154000
14	2.0578000	42	2.5613000	70	2.5002000
15	2.1780000	43	2.5032000	71	2.3651000
16	2.2974000	44	2.5052000	72	2.4982000
17	2.3644000	45	2.5863000	73	2.3832000
18	2.4206000	46	2.3385000	74	2.3453000
19	2.4357000	47	2.1668000	75	2.6093000
20	2.4328000	48	2.2632000	76	2.5954000
21	2.3670000	49	2.1563000	77	2.5728000
22	2.2403000	50	2.1716000	78	2.7036000
23	2.0509000	51	2.3882000	79	2.6495000
24	2.0710000	52	2.3471000	80	2.6431000
25	2.1024000	53	2.4140000	81	2.7871000
26	2.0074000	54	2.3597000	82	2.5168000
27	2.1382000	55	2.3790000	83	2.4505000
28	2.2889000	56	2.5471000	84	2.5473000

The logged airline data of UK of total miles flown for the period January 1964 to December 1970 is given in Table 5.

Identification of ARARMA model

Since the airline data is a non-stationary time series, we transform it into stationary time series using the transformation

$$\tilde{z}_t = z_t - \varphi_\tau z_{t-\tau}$$

The most significant log τ is defined as the value of τ minimizing $E_{rr}(\tau)$ given in equation

$$Err(\tau) = \frac{\sum_{t=\tau+1}^{N} [z_t - \phi_\tau z_{t-\tau}]^2}{\sum_{t=\tau+1}^{N} z_t^{\ 2}}$$

The most significant τ and $\varphi(\tau)$ are estimated as

$$\hat{\tau} = 24,\ \hat{\phi}(\hat{\tau}) = 1.0532$$

After transforming the non-stationary seasonal series into stationary non-seasonal series, the stationary non-seasonal series can be modeled by a whitening filter. The modeling procedure in the time domain is to compute approximate autoregressive schemes

The Parzen's (1982) ARARMA model is written in the form

$$\hat{z}_t = (1-\varphi_\tau B^\tau)\, z_t \text{ and } g_m(B)\ \tilde{z}_t = a_{t-3}$$

where $g_m(B) = 1+\alpha_m(1)\,B + \alpha_m(2)\,B^2 + \ldots + \alpha_m(m)\,B^m$

The least order $\hat{m}$ is found to be 4 and the transfer function $g_m(B)$ is given by

$$g_4(B) = 1+\hat{a}_4(1)B + \hat{a}_4(2)B^2 + \hat{a}_4(3)\,B^3 + \hat{a}_4(4)\,B^4$$

where $\hat{a}_4(1) = -0.3024$

$\hat{a}_4(2) = -0.0506$

$\hat{a}_4(3) = -0.2972$

$\hat{a}_4(4) = 0.2456$

The proposed ARARMA model for the airline data is

$$\tilde{z}_t = (1-\hat{\phi}(24)\,B^{24})\,z_t,\ \hat{\phi}(24) = 1.0332$$

$$g_4(B)\ \tilde{z}_t = a_t$$

where

$$g_4(B) = 1+\hat{a}_4(1)B + \hat{a}_4(2)B^2 + \hat{a}_4(3)\,B^3 + \hat{a}_4(4)\,B^4$$

Cumulative periodogram check

$C(f_j)$'s are found and the plot of $C(f_j)$ against f_j are drawn. The Kolmogorov-Smirnov 5% and 25% probability limits supply a very rough guide to the significance of apparent deviation and fail in this instance to indicate any significant departure from the assumed model. The limit lines are drawn at distances $\pm K_\varepsilon/\sqrt{q}$.

Forecasting

The ARARMA model for the airline data is written in the form for the lead time $t+l$

$$z_{t+l} = -\{\alpha_4(1)\,z_{t+l-1} + \alpha_4(2)\,z_{t+l-2} + \alpha_4(3)\,z_{t+l-3} + \alpha_4(4)\,z_{t+l-4}\}$$
$$+\ \hat{\phi}(24)\,\{z_{t+l-24} + \alpha_4(1)\,z_{t+l-25} + \alpha_4(2)\,z_{t+l-26}$$
$$+\ \alpha_4(3)\,z_{t+l-27} + \alpha_4(4) z_{t+l-28}\} + a_{t+l}$$

The minimum mean square forecast at lead time l and at origin t are given by

$$\hat{z}_t(l) = -[\{\alpha_4(1)\,z_{t+l-1} + \alpha_4(2)\,z_{t+l-2} + \alpha_4(3)\,z_{t+l-3}$$
$$+\ \alpha_4(4)\,z_{t+l-4}\} + \hat{\phi}(24)\,\{z_{t+l-24} + \alpha_4(1)\,z_{t+l-25}$$
$$+\ \alpha_4(2)\,z_{t+l-26} + \alpha_4(3)\,z_{t+l-27} + \alpha_4(4)\,z_{t+l-28}\} + a_{t+l}]$$

Where

$[z_{t+l}]$ is the conditional expectation at time t.

Indian Journal of Science and Technology Vol. 3 No. 9 (Sep 2010) ISSN: 0974- 6846

Fractional ARIMA modeling

Analysis of monthly rainfall data of Coimbatore city (Tamil Nadu, India) from the month of January 1981 to December 1994

The variability of many hydrological time series results ultimately from fluctuation of weather and climate. Long meteorological records are therefore a natural place to look for persistence. The monthly mean rainfall data of Coimbatore city are taken from the month of January 1981 to the month of December 1994 for analysis. It is observed that the city did not experience any amount of rain in some months. The mean $\bar{x}$ of the observed data x_t is subtracted from each data. In fact the resulting data z_t, t = 1,2,... 168 is named as mean rainfall data.

Table 6. Regression estimator in the case of mean rainfall data.

m	d	m	d
5	-0.0125	13	-0.0538
6	-0.043	14	-0.0222
7	-0.0716	15	-0.0192
8	-0.0772	16	-0.0122
9	-0.0616	17	-0.0118
10	-0.0601	18	-0.0126
11	-0.0533	19	-0.0143
12	-0.0607	20	-0.0130

Identification of fractional ARIMA model

It may be seen that the autocorrelations decay very slowly and it indicates the long-term persistence in the series. The value of d is estimated using equation 1.4.1 and it is found to be d = -0.0538 as given in Table 6. The transformed series can be modeled using either methods proposed by Box and Jenkins (1976) or autoregressive scheme proposed by Parzen (1982).

$$d = \frac{-\sum_{i=1}^{m}\left(z_i - \bar{z}\right)\left(w_i - \bar{w}\right)}{\sum_{i=1}^{m}\left(z_i - \bar{z}\right)^2} \quad \ldots \qquad 1.4.1.$$

and d in AN (d, $\left(d, \frac{\pi^2}{6\sum z_i - \bar{z}}\right)$ and n → ∞ and $\frac{\pi^2}{6}$ is the variance of the asymptotic distribution of ε_j

The proposed model (Sekar & Sreenivasan, 1996) is written in the form

$\nabla^d z_t = w_t$, d=-0.0538

$\varphi(B)\, w_t = a_t$

where, $z_t = x_t - \bar{x}$, t=1, 2.... 168

$\varphi(B) = 1 + \hat{\phi}_1 B + \hat{\phi}_2 B^2 + \ldots + \hat{\phi}_{13} B^{13}$

The value of $\hat{\phi}_i$, i = 1, 2, ...13 are given by

$\hat{\phi}_1$ = -0.08928

$\hat{\phi}_2$ = 0.18026

$\hat{\phi}_3$ = -0.01017

$\hat{\phi}_4$ = 0.07184

$\hat{\phi}_5$ = 0.06950

$\hat{\phi}_6$ = -0.15200

$\hat{\phi}_7$ = 0.13000

$\hat{\phi}_8$ = 0.03414

$\hat{\phi}_9$ = 0.07117

$\hat{\phi}_{10}$ = 0.09038

$\hat{\phi}_{11}$ = -0.28334

$\hat{\phi}_{12}$ = -0.14149

$\hat{\phi}_{13}$ = -0.25995

Cumulative periodogram check

The cumulative periodogram is specifically designed for the detection of periodic pattern in a background of white noise. The periodogram of a time series a_t, t = 1,2,... n is given by

$$I(f_j) = 2/n\left[\left(\sum_{t=1}^{n} a_t \cos t\, 2\pi f_j t\right)^2 + \left(\sum_{t=1}^{n} a_t \sin t\, 2\pi f_j t\right)^2\right]$$

where, $f_j = \frac{j}{n}$ is the frequency.

The normalised cumulative periodogram $C(f_j)$ is given by

$C(f_j) = \frac{\Sigma I(f_j)}{ns^2}$, where s^2 is an estimate of σ_a^2

The Kolmogorov-Smirnov 5% and 25% probability limits supply very rough guide to the significance of apparent deviations.

Analysis of Chennai city traffic accidents data

The number of accidents in Chennai city traffic for the period 1987-1997 is taken for analysis. The logged data of Chennai city traffic accidents is given in Table 7. Autoregressive scheme, ARIMA process, fractional ARIMA process, ARARMA models are identified simultaneously after applying the usual stochastic model building procedure.

The models for the Chennai city traffic accidents are listed below:

The autoregressive model is

$\phi(B) z_t = a_t$, where

Indian Journal of Science and Technology Vol. 3 No. 9 (Sep 2010) ISSN: 0974- 6846

Table 7. Chennai city monthly traffic accidents data for the period 1987-1997.

t	Zt	t	Zt	t	Zt	t	Zt
1	6.068426	34	6.126869	67	5.883322	100	5.968708
2	6.109248	35	6.030685	68	6.180017	101	5.955837
3	6.171700	36	6.142037	69	6.285998	102	6.018593
4	6.042633	37	6.285998	70	6.073044	103	6.073044
5	6.200509	38	6.180017	71	6.047372	104	6.100319
6	6.234411	39	6.244167	72	6.238325	105	6.045005
7	6.230482	40	6.298949	73	6.285998	106	6.086775
8	6.216606	41	6.248043	74	6.146329	107	6.030685
9	6.142037	42	6.192362	75	6.238325	108	6.061457
10	6.284134	43	6.073044	76	6.042633	109	6.082219
11	6.104793	44	6.282267	77	6.100319	110	6.028278
12	6.084499	45	6.200509	78	6.086775	111	6.173786
13	6.047372	46	5.991465	79	6.202536	112	6.148468
14	6.059123	47	5.872118	80	6.173786	113	6.093570
15	6.152733	48	6.354370	81	6.084499	114	6.169611
16	6.113682	49	6.059123	82	6.163315	115	6.196444
17	6.282267	50	6.001415	83	6.001415	116	6.129050
18	6.182085	51	6.082219	84	6.070738	117	6.107023
19	6.154858	52	6.109248	85	6.0913100	118	6.107023
20	6.165418	53	6.63315	86	5.071262	119	6.126869
21	6.111467	54	6.059123	87	6.040255	120	6.006353
22	6.175867	55	6.144186	88	5.924256	121	6.059123
23	6.091310	56	6.115892	89	5.998937	122	6.035481
24	6.120297	57	6.137727	90	6.003887	123	6.230482
25	6.075346	58	6.098074	91	6.040255	124	5.976351
26	6.077642	59	6.059123	92	6.006353	125	6.198479
27	6.175867	60	5.894403	93	6.113682	126	6.284134
28	6.165418	61	6.035481	94	5.955837	127	6.398595
29	6.244167	62	5.955837	95	5.820083	128	6.336826
30	6.216606	63	6.082219	96	5.913503	129	6.24107
31	6.118097	64	6.056784	97	5.973810	130	6.122493
32	6.059123	65	6.028278	98	6.003887	131	5.888878
33	6.206576	66	6.059123	99	6.059123	132	5.991460

$\phi(B) = I - \hat{\phi}_1 B \quad \hat{\phi}_1 = 0.9924$

The ARIMA model is

$\nabla z_t = (I - \theta B) a_t,$

Where

$\hat{\theta} = 0.6451$

Fractional ARIMA models is

$\nabla^d z_t = w_t, \quad d = 0.200$

$\phi(B) w_t = a_t$

$\phi(B) = 1 + \hat{\phi}_1 B, \ \hat{\phi}_1 = -0.9910$

ARARMA model is

$\tilde{z}_t = (1 - \hat{\phi}(1)B)\, z_t, \ \hat{\phi}(1) = 0.997$

$g_6(B)\, \tilde{z}_t = a_t$, where

$$g_6(B) = 1 + \hat{\alpha}_6(1) B + \hat{\alpha}_6(2) B^2 + \hat{\alpha}_6(3) B^3 + \hat{\alpha}_6(4) B^4 + \hat{\alpha}_6(5) B^5 + \hat{\alpha}_6(6) B^6$$

and

$\hat{\alpha}_6(1) = 0.52015$

$\hat{\alpha}_6(2) = 0.37378$

$\hat{\alpha}_6(3) = 0.24530$

$\hat{\alpha}_6(4) = 0.107597$

$\hat{\alpha}_6(5) = 0.13438$

$\hat{\alpha}_6(6) = 0.21097$

Among the different models identified, the best model can be chosen by applying the diagnostic check criterion.

References

1. Box GEP and Jenkins GM (1976) Time series analysis: forecasting and control, revised edn, San Francisco: Holden-day.
2. Hosking JRM (1981) Fractional differencing. *Biometrika*, 68(1), 165-176.
3. Montgomary DC and Johnson LA (1976) Forecasting and time series analysis. McGraw Hill Inc., San Francisco. pp: 231-232.
4. Parzen E (1982) ARARMA models for time series analysis and forecasting. *J. Forecasting.* 1, 67-82.
5. Sekar P and Sreenivasan M (1996) Simulation and modeling of time series using fractional differencing. Proc. of Int. Conf. on Stochastic Process. Dec. 26-29, Cochin, India, pp:225-233.