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 (Montgomery & 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_i
ii. The logged series differenced with respect to months only, ∇z_i
iii. The logged series differenced with respect to years only, ∇∇z_i
iv. The logged series differenced with respect to months and years, ∇∇∇z_i
v. The autocorrelation of z_i, ∇z_i, ∇z_i, ∇∇z_i are shown in Table 2.

The autocorrelations for z_i’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 ∇∇∇z_i 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

\[ w_t = (1-θB)(1-θB^{12})a_t, \quad w_t = ∇∇z_t \]

\[ w_{t+12} = a_{t+12} - θ a_{t+11} - Θ a_{t+12} \]

\[ w_{t+k} = a_{t+k} - θ a_{t+k-1} - Θ a_{t+k-12} + θΘa_{t+k-13} \]

The autocovariance at lag k is defined as

\[ γ_k = E [w_t w_{t+k}] \]

\[ γ_0 = E [w_t^2] = (1 + θ^2 + Θ^2)σ^2 \]

\[ γ_1 = (1 + θ^2)σ^2 = -θ \]

\[ γ_11 = θσ^2 \]

\[ γ_12 = (1 + θ^2)σ^2 = -Θ \]

The autocorrelation at lag k is defined as

\[ ρ_k = \frac{γ_k}{γ_0} \]

\[ ρ_1 = \frac{-θ(1+θ^2)σ^2}{(1+θ^2)(1+θ^2)σ^2} = \frac{-θ}{1+θ^2} \] \hspace{1cm} (1.1)

\[ ρ_{12} = \frac{γ_{12}}{γ_0} = \frac{-Θ(1+θ^2)σ^2}{(1+θ^2)(1+θ^2)σ^2} = \frac{-Θ}{1+θ^2} \] \hspace{1cm} (1.2)

It is found that for the airline data

\[ ρ_1 = -0.40, \quad ρ_{12} = -0.339 \]

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

We obtain \[ 0.40 θ^2 - Θ + 0.40 = 0 \]

\[ 0.339 Θ^2 - Θ + 0.339 = 0 \]

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

\[ θ = 0.50, \quad Θ = 0.3907 \]
Where \( \beta \) is a general symbol for \( p+q+P+Q \) parameters.

The unconditional sum of squares function is given by:

\[
\sum_{t=-\infty}^{\infty} \left[ \frac{a_t}{\beta} \right]^2 + \sum_{t=-\infty}^{\infty} \left[ \frac{w_t}{\beta} - \theta \right]^2 \]

conditional on \( \beta \) and \( w \).

| lagk | ACF of \( z_t \) | ACF of \( \nabla z_t \) | ACF of \( \nabla^2 z_t \) | ACF of \( \nabla^4 z_t \) |
|------|-----------------|-----------------|-----------------|-----------------|
| 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 \( n \), \( I(\beta, \sigma_a) \) is dominated by\n
\[
s(\beta) = \frac{s(\beta)}{2\sigma_a^2^2}
\]

The contours of the unconditional sum of square function in the space parameters \( \beta \) are very nearly contours of the log likelihood.

Calculation of the unconditional sum of squares \( s(\beta) \) for the multiplicative model \((0,1,1) \times (0,1,1)_{12}\)

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

\[
w_i = (1-\Theta) F (1-\Theta B)^2_1 \ a_t \text{ or } w_i = (1-\Theta) F (1-\Theta B)'_1 \ a_t
\]

Where \( F \) is the forward shift operator.

Assume that

\[
E(w_i) = \mu = 0
\]

We can write

\[
\{e_i\} = \{w_i\} + \theta [e_i+1] - \Theta [e_i+12] - \Theta \ [e_i+13]
\]

where \( \{w_i\} = \{w_i\} \text{ for } t = 1, 2, ..., n. \)

1. Assume that \( \{e_i\} = 0 \text{ if } t > n \)
2. Knowing the values of \( w_i \), the values of \( \{e_i\} \text{'s are calculated recursively up to } e_i \)
3. The values of \( \{e_i\}, \{e_i\} \text{, ..., } e_{i+12} \text{ are distributed independently of } w_{i+2} \text{,...} \)
4. Knowing all the values of \( e_{12}, e_{11}, ..., e_0 \) We back forecast the values of \( w_{12}, w_{11}, ..., w_0 \text{ for } i > 12. \)
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

\[ \hat{\theta} = 0.77, \quad \hat{\Theta} = 0.84. \]

### 1.2.4. Iterative calculation of least square estimates
For the multiplicative model (0, 1, 1), x (0, 1, 1)_{12}, we have

\[ a_{i,0} = (\theta - \Theta) x_{i,1} + (\Theta - \Theta) x_{i,2} + a_{i}, \]

where \[ x_{i,1} = -\frac{\partial a_{i,0}}{\partial \theta}, \quad x_{i,2} = \frac{\partial a_{i,0}}{\partial \Theta} \]

and \[ \theta, \quad \Theta \] are assigned initial values. The derivatives \[ x_{i,1}, x_{i,2} \] are calculated numerically.

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

| Iteration | \( \theta \) | \( \Theta \) | Sum of squares s(\beta) |
|-----------|-------------|-------------|------------------------|
| 1         | 0.6468      | 0.7209      | 0.750 0.870 0.6119      |
| 2         | 0.6817      | 0.8450      | 0.750 0.880 0.6156      |
| 3         | 0.6897      | 0.8922      | 0.750 0.890 0.6212      |
| 4         | 0.6915      | 0.8836      | 0.750 0.880 0.6127      |
| 5         | 0.6919      | 0.8880      | 0.750 0.870 0.6101      |
| 6         | 0.6920      | 0.8859      | 0.750 0.860 0.6097      |
| 7         | 0.6920      | 0.8869      | 0.770 0.800 0.6122      |
| 8         | 0.6920      | 0.8869      | 0.770 0.800 0.6122      |
| 9         | 0.6920      | 0.8867      | 0.770 0.800 0.6122      |
| 10        | 0.6920      | 0.8866      | 0.770 0.800 0.6110      |
| 11        | 0.6920      | 0.8866      | 0.770 0.800 0.6101      |
| 12        | 0.6920      | 0.8866      | 0.770 0.800 0.6101      |
| 13        | 0.6920      | 0.8866      | 0.770 0.800 0.6101      |
| 14        | 0.6920      | 0.8866      | 0.770 0.800 0.6101      |
| 15        | 0.6920      | 0.8866      | 0.770 0.800 0.6101      |

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 \( C(f) \) against \( f \) would be scattered about a straight line joining the points (0.0) and (0.5, 1). For the model identified using ARIMA multiplicative model \( C(f) \)’s are found and the plot of \( C(f) \) against \( f \) 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 \( \pm K_s / \sqrt{q} \).

The cumulative periodogram are drawn for the models having least square estimates \( \hat{\theta} = 0.692, \quad \hat{\Theta} = 0.8866 \) and \( \hat{\theta} = 0.77, \quad \hat{\Theta} = 0.84 \). The 5% limit lines inserted deviate from the theoretical lines by \( \pm 1.36 / \sqrt{51} \). The 25% limit lines drawn deviate from the theoretical lines by \( \pm 1.02 / \sqrt{51} \). Therefore the least square estimates are taken to be \( \hat{\theta} = 0.692, \quad \hat{\Theta} = 0.8866 \).
The future values of the time series are estimated by minimizing the error function
\[ Err(t) = \frac{1}{t} \sum_{i=t}^{N} [z_i - \phi \bar{z}_{i-t}]^2 \]

The most significant \( \tau \) and \( \phi(\tau) \) 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
\[ \tilde{z}_i = (1 - \phi_i B^i) z_i + \alpha_i(B) \tilde{z}_i = a_{i3} \]
where \( g_m(B) = 1 + \alpha_m(1)B + \alpha_m(2)B^2 + \ldots + \alpha_m(m)B^m \)

The least order \( m \) is found to be 4 and the transfer function \( g_m(B) \) is given by
\[ g_4(B) = 1 + \tilde{a}_4(1)B + \tilde{a}_4(2)B^2 + \tilde{a}_4(3)B^3 + \tilde{a}_4(4)B^4 \]

The proposed ARARMA model for the airline data is
\[ \tilde{z}_i = (1 - \phi(24)B^{24}) z_i, \phi(24) = 1.0332 \]
\[ g_4(B) \tilde{z}_i = a_i \]
where
\[ g_4(B) = 1 + \tilde{a}_4(1)B + \tilde{a}_4(2)B^2 + \tilde{a}_4(3)B^3 + \tilde{a}_4(4)B^4 \]

Cumulative periodogram check
C(f)’s are found and the plot of C(f) against f is 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 \( \pm K_0/\sqrt{q} \).

Forecasting
The ARARMA model for the airline data is written in the form for the lead time \( \tau + 1 \)
\[ z_{t+1} = -\{a_1(1)z_{t+1} + a_4(2)z_{t+2} + a_4(3)z_{t+3} + a_4(4)z_{t+4}\} + \phi(24)\{z_{t+24} + a_4(1)z_{t+25} + a_4(2)z_{t+26} + a_4(3)z_{t+27} + a_4(4)z_{t+28}\} + a_{t+1} \]
The minimum mean square forecast at lead time \( \tau + 1 \) and at origin \( t \) are given by
\[ \tilde{z}_i(\tau + 1) = \{a_1(1)z_{t+1} + a_4(2)z_{t+2} + a_4(3)z_{t+3}\} + \phi(24)\{z_{t+24} + a_4(1)z_{t+25} + a_4(2)z_{t+26} + a_4(3)z_{t+27} + a_4(4)z_{t+28}\} + a_{t+1} \]

Where
\[ [z_{t+1}] \] is the conditional expectation at time \( t \).
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 $x$ of the observed data $x_t$ is subtracted from each data. In fact the resulting data $z_t = x_t - \overline{x}$ is named as mean rainfall data.

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).

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

and $d$ in AN ($d_i$), \(\frac{\pi^2}{6} \sum_{i=1}^{m} (z_i - \overline{z})^2\) and $n \rightarrow \infty$ and $\frac{\pi^2}{6}$ is the variance of the asymptotic distribution of $\varepsilon_i$

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

$$\phi(B) z_t = \sum_{i=1}^{\hat{\phi}} \hat{\phi}_i B^i + \sum_{i=1}^{\hat{\phi}} \hat{\phi}_i B^i \quad \ldots \quad 1.4.1$$

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

| $m$ | $d$ | $m$ | $d$ |
|-----|-----|-----|-----|
| 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 |

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 $a_t, t = 1,2, \ldots n$ is given by

$$I(f_j) = \frac{2}{n} \sum_{i=1}^{n} a_i \cos 2\pi f_j t + \sum_{i=1}^{n} a_i \sin 2\pi f_j t$$

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

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

$$C(f_j) = \frac{\sum I(f_j)}{ns^2}$$

where, $s^2$ is an estimate of $a_s^2$

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

$$\phi(B) z_t = a_t, \quad \text{where}$$
The ARIMA model is
$$\nabla z_t = (1-\theta B) a_t,$$
where
$$\theta = 0.6451$$

Fractional ARIMA models is
$$\nabla^d z_t = w_t,$$  
$$d = 0.200$$
$$\phi (B) w_t = a_t$$
$$\phi (B) = 1 + \tilde{\phi} B, \tilde{\phi} = -0.9910$$

ARARMA model is
$$\tilde{z}_t = (1 - \tilde{\phi} (1)B) z_t, \tilde{\phi} (1) = 0.997$$
$$g_6(B) z_t = a_t,$$  
$$g_6(B) = 1 + \tilde{\alpha}_6 (1) B + \tilde{\alpha}_6 (2) B^2 + \tilde{\alpha}_6 (3) B^3 + \tilde{\alpha}_6 (4) B^4 + \tilde{\alpha}_6 (5) B^5 + \tilde{\alpha}_6 (6) B^6$$

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. Montgomery 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.

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  | 6.171000 | 36 | 6.142037 | 69 | 6.285998 | 102 | 6.018593 |
| 4  | 6.042633 | 37 | 6.285998 | 70 | 6.073043 | 103 | 6.07044 |
| 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 | 6.289894 | 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 | 6.104794 | 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 | 6.591465 | 79 | 6.202535 | 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 | 6.066533 |
| 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 | 6.063535 | 125 | 6.198479 |
| 27 | 6.175867 | 60 | 5.894403 | 93 | 6.116382 | 126 | 6.285348 |
| 28 | 6.165418 | 61 | 6.035481 | 94 | 5.955837 | 127 | 6.398595 |
| 29 | 6.244167 | 62 | 5.955837 | 95 | 5.802348 | 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 |