ISSN: 2456–5474 RNI No.UPBIL/2016/68367 Vol-2* Issue-6* July - 2017
Innovation The Research Concept

Time Series and its Properties

Abstract

In this paper, a time series {$X_t(\omega)$, t Є T} on (Ω , C, P) is explained. Where X is a random variable (r.v.). The properties of time series with supporting real life examples have been taken and conclusion have been drawn by testing methodology of hypothesis. Number of police station and crime data for 33 years from Parbhani District of Maharastra State were analyzed.

Keywords: Time series, regression model, auto-covariance, auto-correlation function.

B. L. Bable
Associate Professor,
Deptt.of Statistics,
Dnyanopasak Mahavidyalaya,
Parbhani

D. D. Pawar
Professor and Research,
Deptt.of Statistics,
N.E.S. Science College,
Nanded

Introduction

Our aim here is to illustrate a few properties of time series with supporting real life examples. Concepts of auto covariance and auto correlation are shown to be useful which can be easily introduced. In this article we have used number of police station and crime data of 1970 to 2002 at Parbhani District to illustrate most of properties theoretically established.

Basic Concepts

Basic definitions and few properties of stationary time series are given in this section.

Definition 2.1

A time series: Let (Ω, C, P) be a probability space let T be an index set. A real valued time series is a real valued function X(t ,ω) defined on T x Ω such that for each fixed t Є T, X(t, ω) is a random variable on (Ω, C, P).

The function X(t, ω) is written as X(ω) or X_t and a time series considered as a collection {X_t : t Є T}, of random variables [9].

Definition 2.2

Stationary Time Series

A process whose probability structure does not change with time is called stationary. Broadly speaking a time series is said to be stationary, if there is no systematic change in mean i.e. no trend and there is no systematic change in variance.

Definition 2.3

Strictly Stationary Time Series

A time series is called strictly stationary, if their joint distribution function satisfies.

$$F_{X_{t1} X_{t2} \cdots X_{tn}}(X_{t1}\, X_{t2} \cdots X_{tn}) = F_{X_{t1+h} X_{t2+h} \cdots X_{tn+h}}(X_{t1}\, X_{2t} \cdots X_{tn}) \quad \ldots (1)$$

For example, let I = [-T, T] be a set of integers. Let I_1= [t_1, t_2 ... t_n] and I_2 = [t_1+h, t_2+h ... t_{n+h}] be two subset of I, where h is an integer. A random process X_t defined over I_1 has n random variables has a joint density function f(X_t for t Є I_1). A time series is strictly stationary if f(X_t for t Є I_1) = f(X_t for t Є I_2) for every n, every h, and every subset t_1, t_2, ...t_n.

Main Results

Theorem 3.1: If {X_t: t Є T}, is strictly stationary with E{ | X_t | } < α and

E{ |X_t – μ | } < β then,

$E(X_t) = E(X_{t+h})$, for all t, h} … (2)

$E[(X_{t1} – \mu)(X_{t2} – \mu)] = E[(X_{t1+h} – \mu)(X_{t2+h} – \mu)]$, for all t_1, t_2, h

Proof

Proof follows from definition (2.3)

In usual cases above equation (2) is used to determine that a time series is stationary i.e. there is no trend.

Definition 3.1

Weakly Stationary Time Series

A time series is called weakly stationary if

1. The expected value of X_t is a constant for all t.
2. The covariance matrix of (X_{t1}, X_{t2}, X_{tn}) is same as covariance matrix of (X_{t1+h}, X_{t2+h}, …. X_{tn+h}).

A look in the covariance matrix (X_{t1} X_{t2} … X_{tn}) would show that diagonal terms would contain terms covariance (X_{ti}, X_{ti}) which are essentially variances and off diagonal terms would contains terms like covariance (X_{ti}, X_{tj}). Hence, the definitions to follow assume importance. Since these involve elements from the same set {X_{ti}}, the variances and co- variances are called auto- variances and auto-co variances.

Definition 3.2

Auto-Covariance Function

The covariance between {X_t} and {X_{t+h}} separated by h time unit is called auto-covariance at lag h and is denoted by Υ (h).

$\Upsilon(h) = Cov(X_t, X_{t+h}) = E\{X_t - \mu\}\{X_{t+h} - \mu\}$ … (3)

the function Υ (h) is called the auto covariance function.

Definition 3.3

The Auto Correlation Function

The correlation between observations which are separated by h time unit is called auto-correlation at lag h. It is given by

$$P(h) = \frac{E\{X_t - \mu\}\{X_{t+h} - \mu\}}{[E\{X_t - \mu\}^2 E\{X_{t+h} - \mu\}^2]^{1/2}} \quad \text{… (4)}$$

$$= \frac{\Upsilon(h)}{[E\{X_t - \mu\}^2 E\{X_{t+h} - \mu\}^2]^{1/2}}$$

where μ is mean.

Remark 3.1

For a stationary time series the variance at time (t + h) is same as that at time t, **[2, 3, 4, 5, 6]**. Thus, the auto correlation at lag h is

$$\rho(h) = \frac{\Upsilon(h)}{\Upsilon(0)} \quad \text{… (5)}$$

Remark 3.2

For h = 0, we get, ρ (0) = 1.

For application, attempts have been made to establish that stationary satisfy equation (1) and (5).

Theorem 3.2

The covariance of a real valued stationary time series is an even function of h.

i. e. $\Upsilon(h) = \Upsilon(-h)$.

Proof

We assume that without loss of generality, E{X_t} = 0, then since the series is stationary we get , $E\{X_t X_{t+h}\} = \Upsilon(h)$, for all t and t + h contained in the index set. Therefore if we set $t_0 = t_1 - h$,

$\Upsilon(h) = E\{X_{t0} X_{t0+h}\} = E\{X_{t1} X_{t1-h}\} = \Upsilon(-h)$ … (6)

Proved.

Theorem 3.3

Let X_t's be independently and identically distributed with $E(X_t) = \mu$

And $var(X_t) = \sigma^2$

Then

$$\Upsilon(t, h) = E(X_t, X_h) = \sigma^2, \; t = h$$
$$= 0, \; t \neq h$$

This process is stationary in the strict sense.

Testing Procedure

Inference Concerning Slope (β_1)

We set up null hypothesis for testing H_0: β_1 = 0 Vs H_1: $\beta_1 > 0$ for α = 0.05 percent level using t distribution with degrees of freedom(d. f.) is equal to n – 2 were considered. Whether the hypothesis H_0 is significant or not for number of police station series and for number of crime series by using corresponding d. f. for t and h i.e. 31 and 19 d. f. respectively.

$t_{n-2} = b - \beta_1 / s_e$

Where β_1 is the slope of the regression line. s_e is the standard error.

4.2: Example of Time Series

Number of police station and crime data were collected from Vasantrao Naik Marathwada Agricultural University, Parbhani [1]. Hence we have Table 5.1, two dimensional time series t_i, i = 1, 2 to the district Parbhani. Table 5.1A and Table 5.1B shows the results of descriptive statistics, Table 5.1C and Table 5.2C shows linear trend analysis. All the linear trends were found to be not significant in the series of police station, but in crime case trends were found therefore there is a significant trend.

Over the years many scientists have analyzed rainfall, temperature, humidity, agricultural area, production and productivity of region of Maharastra state, [2, 3, 4, 5, 6, 8]. Most of them have treated the time series for each of the revenue districts as independent time series and tried to examine the stability or non-stability depending upon series. Most of the times non- stability has been concluded and hence possibly any sort of different treatment was possibility never thought of. In this investigation we treat the series first and individual series. The method of testing intercept (β_0 = 0) and regression coefficient (β_1 = 0), Hooda R.P. [10] and for testing correlation coefficient Bhattacharya G.K. and Johanson R.A. [7].

The regression analysis tool provided in MS-Excel was used to compute β_0, β_1, corresponding SE, t-values for the coefficients in regression models. Results are reported in Table 5.1C and Table5.2C. Elementary statistical analysis is reported in Table5.1A and Table 5.1B.

Table5.1C shows that the model,

$$X_t = \beta_0 + \beta_1 t + \epsilon,$$

When applied to the data indicates H_0: β_1 = 0 is true. Hence X_t is a not having trend for series first of the district.

Where,

1. X_t are the number of police station and crime data series.
2. t is the time (years) variable.
3. ϵ is a random error term normally distributed with mean 0 and variance σ^2.

Number of police station and crime data series X_t is the dependent variable and time t in (years) is the independent variable.

Values of auto covariance computed for various values of h are given in Table5.2A. Number of police station and crime data values for district was input as a matrix to the software. Defining

$A = x_1, x_2 \dots x_{n-h}$

$B = x_{h+0}, x_{h+1} \dots x_n$

Υ (h) = cov (A, B) were computed for various values of h. Since the time series constituted of 33 values, at least 10 values were included in the computation.The relation between Υ (h) were examined using model, Table-5.2C.

$$\Upsilon(h) = \beta_0 + \beta_1 h + \in,$$

The testing shows that, both the hypothesis $\beta_0 = 0$ and $\beta_1 = 0$ test is not positive. Table5.2C was obtained by regressing values of Υ (h) and h, using "Data Analysis Tools" provided in MS Excel. Table 5.2A formed the input for table 5.2C. In other wards, Υ(h) are all zero in the number of police station data series of the district, trends were not found showing that X_t, X_{t+h} are not dependent in first series of the district and there is no trend in that series.

Conclusion

It was observed that t values are therefore not significant for number of police station series of the district but found significant for number of crime, i.e. conclude that X_t depend on t for number of crime series of the district [7]. Similarly, Υ_{ij}(h) depend on h to mean that, 'linear relation' rather than 'relation' Υ_{ij}(h) and h for crime series. The testing shows that, for the hypothesis $\beta_1 = 0$, there is positive evedence for crime series of the district.

Generally it is expected, police station and crime series data (annual) over a long period at any region to be not stationary time series. The results is not conform with the series of number of police station but conform with the series of crime in the district.

5.1. Police Station and Crime Time Series Data Treated as Scalar Time Series

Table 5.1 contains the results for scalar series approach.

The model considered was:

$$X_i(t) = (\beta_0)_i + (\beta_1)_i t + \in_i(t), \quad i = 1, 2. \quad \dots (7)$$

Where X_i is number of police station series or crime series in the district, t is the time series variable, β_0 = the intercept, β_1 = the slope, $\in_i$ is the random error. Number of police station series or crime X_i is the dependent variable and time t in years is the independent variable.

Descriptive Statistics

Table-5.1: Police Station and Crime Series Data of Parbhani District.

Sr. No.	District→ Years ↓	police station	crime
1	1970	27	2630
2	1971	27	2948
3	1972	27	3143
4	1973	27	1969
5	1974	27	6180
6	1975	29	7170
7	1976	29	1669
8	1977	29	2366
9	1978	29	2574
10	1979	29	4720
11	1980	29	5724
12	1981	29	5654
13	1982	29	5724
14	1983	29	6714
15	1884	29	4274
16	1985	29	5650
17	1986	26	5941
18	1987	26	5951
19	1988	26	5940
20	1989	26	5931
21	1990	26	5442
22	1991	26	5404
23	1992	27	4529
24	1993	27	6426
25	1994	27	6912
26	1985	27	6518
27	1996	27	6521
28	1997	28	6627
29	1998	28	6921
30	1999	28	6527
31	2000	28	4051
32	2001	29	4125
33	2002	29	4136

Table-5.1 A

Elementary Statistics of Observed Minimum, Maximum, Mean, Standard Deviation and Coefficient of Variation of Police Station in Parbhani District.

Parbhani District	Minimum	Maximum	Mean	S.D.	C.V.
Police station	26	29	27.73	1.16	4.19

Table-5.1B

Elementary Statistics of Observed Minimum, Maximum, Mean, Standard Deviation and Coefficient of Variation of Crime in Parbhani District.

Parbhani	Minimum	Maximum	Mean	S.D.	C.V.
crime	1669	7170	5060.94	2756.01	54.46

Table-5.1C

Linear Regression Analysis of Data to Determine Trend Eq (7)

Parbhani District	Coefficients		Standard Error	t Stat	Significance	P Value	Lower 95%	Upper 95%
Police station	β_0	27.93	0.43	65.70	S	0.00	27.06	28.79
	β_1	-0.01	0.02	-0.54	NS	0.60	-0.06	0.03
crime	β_0	3692.16	511.49	7.22	S	0.00	2648.95	4735.36
	β_1	80.52	26.25	3.07	S	0.00	26.98	134.06

t =2.04 is the critical value for 31 d f at 5% L. S. * shows the significant value

A look at the table 5.1A &B shows that all of them have similar values of CV. Which indicates that their dispersion is almost identical? Trends were found to be not significant in number of police station series of the district but significant in number of crime series. In absence of linear trend, with reasonably low CV values can be taken as evidence of series being stationary series individually in number of police station series of the district.

Further search for evidences of stability included determination of auto covariance and their dependency on lag variable h (Table 5.2A). Such an analysis requires an assumption of AR(Auto-regressive) model **[9]** Eq(8). Therefore a real test for stationary property of the time series can come by way of establishing auto- covariance's which do not depend on the lag variable

$$X_t = C + \Phi X_{t-h} + \epsilon_t, \quad h = 0, 1, 2, \ldots 20 \ldots \quad (8)$$

Table-5.2 A

Auto Variances: Individual Column Treated as Ordinary Time Series for lag values (h = 0, 1, 2 … 15) about both the data.

Parbhani District		
lag h	**Number of police station**	**Crime**
0	1.35	2524608.12
1	1.11	1293406.17
2	0.86	423024.31
3	0.63	302126.46
4	0.38	144324.03
5	0.09	526243.99
6	-0.22	795584.13
7	-0.40	448192.90
8	-0.58	656125.50
9	-0.78	-43586.70
10	-1.03	-183134.18
11	-1.31	121123.84
12	-1.10	52968.76
13	-0.86	78132.20
14	-0.58	308479.01
15	-0.58	308479.01
16	-0.25	-8874.49
17	0.33	-92326.43
18	0.74	-722677.11
19	0.71	-590445.73
20	0.68	-285895.29

Table-5.2 B

Correlation Coefficient between h and Auto Covariance is

Corr. Coefficient	-0.31	-0.66*

*Correlation coefficient r = 0.433 is the critical value for 19 d f at 5% L. S. *shows the significant value.*

Correlation's between Υ_{ij} **(h)** and **h** were found not ***significant*** in number of police stations but ***significant*** in number of crime series of the district only showing that the time series can be reasonably assumed to be **stationary** i.e. not having trend for number of police stations and trend was found for number crime data series**.**

Table-5.2 C Linear Regression Analysis of Lag Values h vs Covariance.

District	**Coefficients**		**Standard Error**	**t Stat**	**Significane**	**P Value**	**Lower 95%**	**Upper 95%**
Police station	β_0	0.32	0.30	1.05	NS	0.31	-0.32	0.96
	β_1	-0.04	0.03	-1.41	NS	0.18	-0.10	0.02
Crime	β_0	955122.72	206983.81	4.61	S	0.00	521900.48	1388344.95
	β_1	-72173.70	18666.72	-3.87	S	0.00	-111243.60	-33103.81

t =2.1 is the critical value for 19 d f at 5% L. S. * shows the significant value

References

1. *Anonymous. Vasantrao Naik Marathwada Agricultural University, Parbhani (1981-2004).*
2. *Bable, B. L. and Acharya H. S., Time Series: The Concept and Properties to the Stationary Time Series; Bulletin of the Marathwada Mathematical Society., 7(1), (2006), 1-10.*
3. *Bable, B. L. and Pawar D. D., Trends of District-wise Scalar Time Series; International Journal of Statistica and Mathematica, ISSN 1(1), (2011), 01-07.*
4. *Bable, B. L. and Pawar D. D., Formation of Trends in Scalar Time Series; International Research Jour Internat. Res. J. of Agric. Eco. And Stat., 3(1) (2014), 111-115.*
5. *Bable, B. L. and Pawar D. D., Properties of Weak and Strict Stationary Time Series; International Research Jour. of Asian Resonance, E:ISSN No.2349-9443) (2015), 55-59.*
6. *Bable, B. L. and Pawar D. D., Trends in Scalar Time Series; International Journal of Periodic Research E:ISSN No.2349-9435) (2015), 01-06.*
7. *Bhattacharya, G. K. and Johanson R. A., Statistical Concepts and Methods. John Wiley and Sons, New York (1977).*
8. *Das, P. K., Statistics and Its Applications to Meteorology. Indian J. pure appl. Math., 26(6) (1995) , 531-546.*
9. *Fuller W. A., Introduction to Statistical Time Series. John Wiley and Sons, New York (1976).*
10. *Hooda R. P., Statistics for Business and Economics. Machmillan India Ltd., New Delhi (2003).*