Trends of District wise Scalar Time Series

B. L. Bable [1] , D. D. Pawar [2]

[1]Head , Department of Statistics, Dnyanopasak Mahavidyalaya,
Parbhani - 431 401, Maharastra, INDIA .
Ph. : 02452-240888 , Mobile : 09420817528

[2]Associate Professor and Chairman of BOS in Statistics,
N.E.S. Science College , Nanded
Maharastra , INDIA

Corresponding Adresses

Email ID : b.bable@rediffmail.com

Research Article

Abstract: In this paper, a time series { X(t, ω) , t Є T } where X is a random variable (r .v.) on (Ω , C, P) is explained. The properties of stability with supporting real life examples have been taken and conclusions have been drawn by testing methodology of hypothesis. Temperature data for 33 years from five districts of Marathwada of Maharastra State were analyzed .

A preliminary discussion of properties of time series precedes the actual application to regional district-wise temperature data.

Keywords: Time series , regression equation , auto-covariance , auto-correlation.

1.INTRODUCTION:

Our aim here is to illustrate a few properties of stationary 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 temperature data of 1970 to 2002 at five locations in Marathwada region to illustrate most of properties theoretically established.

2.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 **[8]**.

Definition 2.3: 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 2.4: Strictly stationary time series : A time series is called strictly stationary, if their joint distribution function satisfy

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

Where, the equality must hold for all possible sets of indices ti and (ti + h) in the index set. Further the joint distribution depends only on the distance h between the elements in the index set and not on their actual values.

Theorm 2.1: If $\{ X_t : t \in T \}$, is strictly stationary with $E\{ |X_t| \} < \alpha$ and $E\{ |X_t - \mu| \} < \beta$ then ,

$$E(X_t) = E(X_{t+h}) \text{, for all } t, h$$

and

$$E [(X_{t_1} - \mu)(X_{t_2} - \mu)] = E [(X_{t_{1+h}} - \mu)(X_{t_{2+h}} - \mu)] \text{, for all } t_1, t_2, h \quad \text{….(2)}$$

Proof: Proof follows from definition (2.4).

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

Definition 2.5: 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 $(X_{t1}, X_{t2}, \ldots X_{tn})$ is same as covariance matrix of $(X_{t1+h}, X_{t2+h}, \ldots X_{tn+h})$.

A look in the covariance matrix $(X_{t1}\ X_{t2} \ldots X_{tn})$ 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 2.6: 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 $\Upsilon(h)$.

$$\Upsilon (h) = \mathrm{cov} (X_t , X_{t+h}) = E\{ X_t - \mu \}\{ X_{t+h} - \mu \} \quad \text{….(3)}$$

the function $\Upsilon(h)$ is called the auto covariance function.

Definition 2.7: The auto correlation function: The correlation between observation 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 2.1: For a vector stationary time series the variance at time (it + h) is same as that at time it. Thus, the auto correlation at lag h is

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

Remark 2.2: For h = 0, we get , $\rho(0) = 1$.

For application attempts have been made to establish that temperature at certain districts of Marathwada satisfy equation (1) and (5).

Theorem 2.2: 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 $t_0 = t_1 - h$,

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

proved.

Theorem 2.3: Let X_t's be independently and identically distributed with $E(X_t) = \mu$ and $\mathrm{var}(X_t) = \sigma^2$ then

$$\Upsilon (t, k) = E(X_t , X_k) = \sigma^2, \quad t = k$$
$$= 0, \quad t \neq k$$

This process is stationary in the strict sense.

3. TESTING PROCEDURE

3.1: Inference concerning slope (β_1): For testing $H_0 : \beta_1 = 0$ Vs $H_1: \beta_1 > 0$ for $\alpha = 0.05$ percent level using t distribution with degrees of freedom is equal to n – 2 were considered.

$t_{n-2} = \beta_1 / s_{\beta 1}$

where β_1 is the slope of the regression line and $s_{\beta 1} = s_e / s_t$ and $s_e = [\, SSE \;/ n - 2]^{1/2}$, sum of squares due to errors (SSE) = $(s_t^{\,2} - s_{tx}^{\,2} / s_t^{\,2})$, $s_{tx} = \Sigma(t_i - \bar{t})(X_t - \bar{X})$ $s_t^2 = \Sigma(t_i - \bar{t})^2$; $s_x^{\,2} = \Sigma(X_t - \bar{X})^2$ .Where SSE is the sum of squares due to error or residual sum of squares .

3.2: Example of time series: temperature data of Marathwada region were collected from five districts namely Aurangabad, Parbhani, Beed, Osmanabad, and Nanded. The data were collected from Socio Economic Review and District Statistical Abstract, Directorate Economics and Statistics Government of Maharastra Bombay, Maharastra Quarterly Bulletin of Economics and Statistics, Directorate of Economics and Statistics Government of Maharastra , Bombay and Hand Book of Basic Statistics of Maharastra State **[2, 3 , 4]**. Hence. we have five dimensional time series t_i , i = 1, 2 , 3 , 4 , 5 corresponding to the districts Aurangabad , Parbhani , Beed , Osmanabad and Nanded respectively. Table 4.1A, shows the results of descriptive statistics, table 4.1B and table 4.2C shows linear trend analysis . All the linear trends were found to be not significant except Osmanabad district.

Over the years many scientists have analyzed rainfall , temperature , humidity ,agricultural area , production and productivity of region of Maharastra state, **[1 , 5 , 7 , 10 , 11]** . 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 ($\beta_0 = 0$) and regression coefficient ($\beta_1 = 0$), Hooda R.P. **[9]** and for testing correlation coefficient Bhattacharya G.K. and Johanson R.A**. [6]**. We set up null hypothesis for test statistic used to test ,$H_0 : \beta_1 = 0$ and $H_1 : \beta 1 > 0$, for $\alpha = 0.05$,

$$t_{n-2} = \beta \sqrt{S_{xx}} / \sigma^\wedge \text{, where } \sigma^\wedge = \sqrt{SSE / n-2}$$

The hypothesis H_0 is not significant for both the values of t for 31and 18 d. f. for each districts.

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 4.1B and table4.2C. Elementary statistical analysis is reported in table–4.1A . It is evident from the values of CV that there is hardly any scatter of values around the mean indicating that all the series are not having trend.

Table4.1B shows that the model,

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

When applied to the data indicates $H_0 : \beta_1 = 0$ is true . Hence X_t is a not having trend for four districts except Osmanabad districts .

$X = \beta_0 + \beta_1 t + \in,$

where,

X_t are the annual temperature series .

t is the time (years) variable.

ε is a random error term normally distributed as mean 0 and variance σ^2 .

Temperature X_t in (C^0) 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 table-4.2A ..Temperature values for different districts were input as a matrix to the software. Defining

$A = y_1 , y_2 \ldots\ldots y_{n-h}$

$B = y_{h+0} , y_{h+1} \ldots\ldots y_n$

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-4.2C.

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

the testing shows that , both the hypothesis $\beta_0 = 0$ and $\beta_1 = 0$ test is positive . Table-4.2C was obtained by regressing values of Υ (h) and h , using “Data Analysis Tools” provided in **MS Excel**. Table 4.2A formed the input for table 4.2C. In other wards, Υ (h) are all zero except Osmanbad districts, in the temperature series of Osmanabad district trend was found showing that X_t , X_{t+h} are dependent in temperature series of Osmanabad district and there is a trend in that series . Hence in Osmanabad district temperature X_t is not stationary it presents a trend .

4.Conclusion

It was observed that t values are therefore not significant for the 4 districts , except Osmanabad district i.e. concluded that X_t does not depend on t for 4 districts [5] . Similarly, $\Upsilon_{ij}(h)$ does not depend on h to mean that , 'no linear relation' rather than 'no relation'. The testing shows that, for the hypothesis $\beta_1 = 0$, test is positive for t and h for 4 districts except Osmanabad .

Generally it is expected, temperature (annual) over a long period at any region to be stationary time series. These results does not conform with the series in Osmanbad district i,e. in Osmanabad district trend is found in temperature series.

4.1: Analysis: Temperature

The same strategy of analyzing first individual time series as scalar series and then treating the vector series as the regional time series has been adapted here for maximum temperature.

4.2. Temperature time series treated as scalar time series

Table 4.1 contains the results for scalar series approach.

The model considered was :

$$X_i(t) = (\beta_0)_i + (\beta_1)_i t + \epsilon_i(t), \quad i = 1, 2, \ldots 5 \qquad \text{---- (7)}$$

Where X_i is the annual rainfall series , t is the time seies variable , β_0 = the intercept , β_1 = the slope , ϵ_i is the random error. Rainfall X_i is the dependent variable and time t in years is the independent variable.

Table-4.1: Temperature data of five stations(districts) in Marathwada region

Sr. No.	Districts→ Years ↓	Aurangabad	Parbhani	Osmanabad	Beed	Nanded
1	1970	32.4	41.6	40.9	33	41.9
2	1971	42	43.2	42.7	42.4	43.1
3	1972	42.8	43.7	40.6	43.4	44
4	1973	42.8	43.7	40.4	43.4	44
5	1974	41.1	41.9	41	44.6	44.2
6	1975	42.2	44.2	43	43	43.8
7	1976	42.2	44.2	38.8	39.7	43.8
8	1977	39	40.7	37.2	39.4	40.6
9	1978	39	40.7	37.2	39.4	40.6
10	1979	40.1	42.8	38.7	40.4	42.8
11	1980	41.8	42.4	38.7	41.1	42.8
12	1981	39	41.1	39.5	41.4	41.3
13	1982	40.3	43.3	36.8	38.8	43.6
14	1983	40.6	45	40	44.6	44.6
15	1984	43.6	45.3	39.8	42.4	45.5
16	1985	40.6	42	38.8	42	42.2
17	1986	43.6	45.5	39.3	42.8	42.9
18	1987	40.6	43.8	40.7	46	43.9
19	1988	40.7	43.4	42.9	40.2	43.3
20	1989	40.1	43.8	43.2	40.3	43.9
21	1990	43.4	40.5	43.4	44.3	46
22	1991	40	46.3	43	43.6	45.4
23	1992	38.8	43.8	41.4	41.5	43.9
24	1993	40.4	42.6	41	41.7	40.2
25	1994	41	42.5	40.4	40.9	42.8
26	1995	39	43.2	40	40.3	42.8
27	1996	39.8	39	41	40.4	43.1
28	1997	40.1	39	41	40.4	43.4
29	1998	39.5	44.8	41.9	39.6	43.5

30	1999	40.1	42.3	43.1	39.6	41.6
31	2000	40	42.5	42.2	39.6	41.3
32	2001	39.2	41.6	43.1	40.7	41.6
33	2002	39.5	41.4	42.5	41	41.9

Table-4.1A: Elementary statistics of temperature data (in degree centigrade C^0) of Marathwada region for 33 years (1970-2002).

Cities:	Aurangabad	Parbhani	Osmanabad	Beed	Nanded
Mean:	**40.46**	**42.78**	**40.73**	**41.27**	**43.04**
S.D.:	2.00	1.71	1.85	2.30	1.39
C.V.:	**3.49**	**3.06**	**3.34**	**3.68**	**2.74**

Table-4.1B: Linear regression analysis of temperature data to determine trend Eq(7)

District	Coefficients		Standard Error	t Stat	Significance
Aurangabad	β_0	40.80	0.73	55.87	S
	$\boldsymbol{\beta_1}$	-0.02	0.04	-0.52	NS
Parbhani	β_0	43.26	0.62	69.53	S
	$\boldsymbol{\beta_1}$	-0.03	0.03	-0.87	NS
Osmanabad	β_0	39.34	0.62	63.91	S
	$\boldsymbol{\beta_1}$	0.08	0.03	**2.59***	**S**
Beed	β_0	41.44	0.85	49.00	S
	$\boldsymbol{\beta_1}$	-0.01	0.04	-0.23	NS
Nanded	β_0	43.38	0.51	85.68	S
	$\boldsymbol{\beta_1}$	-0.02	0.03	-0.77	NS

*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 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 4 districts but ***significant*** in ***Osmanabad*** district only. A simple look at the mean values shows that a classification as

C1 = {Aurangabad , Osmanabad }

C2 = {Nanded , Parbhani, Beed }

could be quite feasible.

In absence of linear trend , with reasonably low CV values can be taken as evidence of series being stationary series individually in four districts.

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 **[8]** 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

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

Table-4.2A : Auto variances: Individual column treated as ordinary time series for lag values (h = 0 , 1 , 2 ,…20) about temperature data.

lag h	Aurangabad	Parbhani	Osmanabad	Beed	Nanded
0	4.0	2.9	3.4	5.3	1.9
1	0.1	0.2	2.4	1.0	0.8
2	-0.3	-0.3	1.6	-0.2	-0.1
3	-0.4	0.5	1.1	0.1	-0.3
4	0.7	0.3	0.8	-0.5	-0.2
5	-1.0	-0.4	0.2	-1.5	-0.5
6	-0.6	-0.7	-0.4	0.2	-0.2
7	0.2	0.6	-0.7	0.2	0.6

8	0.2	-0.1	-0.1	-0.4	0.6
9	-0.3	-0.2	0.0	-0.6	-0.6
10	0.0	-1.1	-0.2	-0.5	-0.8
11	0.8	-0.1	-0.8	-0.4	-0.5
12	-0.1	0.2	-0.6	1.1	-0.5
13	0.2	-1.2	-0.2	-1.5	-0.8
14	-0.6	-1.1	-0.2	-1.0	-0.2
15	0.3	0.4	0.2	-0.5	0.6
16	-1.9	0.7	0.5	-0.1	0.5
17	0.1	-0.5	0.7	-1.7	-0.1
18	0.2	0.0	0.8	1.8	0.0
19	0.2	1.1	0.2	1.7	-0.1
20	-2.2	0.3	-0.4	-1.3	-0.4

Table-4.2B : Correlation coefficient between h and Auto covariance is:

Corr. Coefficient	-0.426	-0.207	**-0.526***	-0.292	-0.333

*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 ***significant*** in ***Osmanabad*** district only showing that the time series can be reasonably assumed to be **not stationary** i.e. having trend**.** The coefficient is significant , with negative value showing that Osmanabad has been experiencing significantly declining temperature over the past years .

Table-4.2C : Linear regression analysis of lag values vs covariance.

District	Coefficients		Standard Error	t Stat	Significance
Aurangabad	β_0	0.78	0.46	1.70	NS
	$\boldsymbol{\beta_1}$	-0.08	0.04	-2.05	NS
Parbhani	β_0	0.38	0.38	1.00	NS
	$\boldsymbol{\beta_1}$	-0.03	0.03	-0.92	NS
Osmanabad	β_0	1.28	0.38	3.35	S
	$\boldsymbol{\beta_1}$	-0.09	0.03	**-2.70***	**S**
Beed	β_0	0.79	0.64	1.23	NS
	$\boldsymbol{\beta_1}$	-0.07	0.05	-1.33	NS
Nanded	β_0	0.34	0.27	1.27	NS
	$\boldsymbol{\beta_1}$	-0.04	0.02	-1.54	NS

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

ACKNOWLEDGEMENT:

The authors are thankful to Principal Dr. Kalamse, Science College Nanded , for providing the necessary facilities for the present investigation and encouragement . We are also grateful to Prof. H.S. Acharya, Puna College, PUNE, for his valuable guidance.

REFERENCES:

[1] Acharya , H. S., Quantification of Climatic Component of Food Grains Production; J. of Maharastra Agric. Univ., 16(3): 324-326, 1991

[2] Anonymous , "Socio Economic Review and District Statistical Abstract" , Directorate of Economic and Statistics , Government of Maharastra, Bombay 1970-1997

[3] Anonymous , "Maharastra Quarterly Bulletin of Economics and Statistics", Directorate of Economics and Statistics Govt. of Maharastra , Bombay, 1998.

[4] Anonymous, "Hand Book of Basic Statistics ", Maharastra State 1999-2002.

[5] 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): 1-10, 2006

[6] Bhattacharya , G. K. and Johanson R. A. , Statistical Concepts and Methods. John Wiley and Sons, New York 1977.

[7] Das , P. K., Statistics and Its Applications to Meteorology. Indian J. pure appl. Math., 26(6): 531-546, 1995

[8] Fuller W. A. , Introduction to Statistical Time Series. John Wiley and Sons, New York 1976.

[9] Hooda R. P., Statistics for Business and Economics. Machmillan India Ltd., New Delhi 2003.

[10] Patil , V . D . and Kalyanker S. P ., Impact of Research on Productivity of Agricultural in Maharastra . Agri . Situ . India ., 47(5): 321-326, 1982.

[11] Waghmare , P . R . and Katapallewar B. N . , Districtwise trend in area , production and productivity of oilseeds in Maharastra State , 1980-81 to 1989-90. Indian J. Agric. Econ. , 48(3): 425-426, 1993