P: ISSN NO.: 2321-290X RNI : UPBIL/2013/55327 VOL-5* ISSUE-11* (Supplementary Issue) July- 2018
E: ISSN NO.: 2349-980X **Shrinkhla Ek Shodhparak Vaicharik Patrika**

Districtwise Covariance Scalar Time Series

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. Production of kharif jawar data for 27 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 production of kharif jawar data.

Keywords: Time Series, Regression Equation, Auto-Covariance, Auto-Correlation.

B. L. Bable
Head,
Deptt. of Statistics,
Dnyanopasak Mahavidyalaya,
Parbhani, Maharastra

D. D. Pawar
Professor and Chairman of BOS,
Deptt. of Statistics,
N.E.S. Science College,
Nanded, Maharastra

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 production data of 1976 to 2002 at five locations in Marathwada region to illustrate most of properties theoretically established.

Objectives of the Study

1. To develop theory of time series. Specially improving the theorems which characterize time series.
2. To develop algorithms for analyzing time series, which use the characterizing theorems.
3. By using data from Marathwada Region for validating the algorithms, and testing the methods.
4. To interpret the results of characterizations, in real economic and social terms.

The main purpose of this work is to summarize the research work carried out on the above given objectives and to draw useful conclusions on the basis of auto regressive time series analysis. A way to check trends and randomness in the data scalar time series by using properties of auto covariance.

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 **[12]**.

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,

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

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

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 $\Upsilon(h)$.

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

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

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

Where μ is mean .

Remark 2.1

For a stationary time series the variance at time (it + h) is same as that at time it. Thus, the auto correlation at lag h is

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

Remark 2.2

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

For application attempts have been made to establish that production 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 $t_0 = t_1 - h$,

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

Testing Procedure

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.

Example of Time Series

Production of kharif jawar 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 [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 significant except Parbhani and Beed district.

Over the years many scientists have analyzed rainfall, temperature, humidity, agricultural area, production and productivity of region of Maharastra state, [1, 5, 6, 7, 8, 9, 11, 14, 15]. 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. **[13]** and for testing correlation coefficient Bhattacharya G.K. and Johanson R.A**. [10]**. 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 25 and 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.

Table4.1 B 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 having trend for three districts except Parbhani and Beed districts.

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

Where,

1. X_t are the annual production series.
2. t is the time (years) variable.
3. ε is a random error term normally distributed as mean 0 and variance σ^2 .

Production 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 table-4.2A. Production 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 27 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 production series of Aurngabad, Osmanabad and Nanded districts trend were found showing that X_t, X_{t+h} are dependent in production series of Aurngabad, Osmanabad and Nanded district and there is a trend in that series. Hence in Aurngabad, Osmanabad and Nanded districts X_t is not stationary it presents a trend.

Conclusion

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

Generally it is expected, production of kharif jawar over a long period at any region to be not stationary time series. These results does not conform with the series in Parbhani district i,e. in Parbhani district trend was not found in production of kharif jawar series. But the rest of the four series have trend.

Analysis

Production Time Series

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 production series.

Production Time Series Treated As Scalar Time Series

The model considered was:

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

Where X_i is the production of kharif jawar series, t is the time seies variable, β_0 = the intercept , β_1 = the slope , ϵ_i is the random error. production X_i is the dependent variable and time t in years is the independent variable.

Table-4.1: Districtwise Production of Kharif Jawar of Hundred Hectores of Kharif Jawar Time Series Data of Five Stations (Districts) In Marathwada Region

Sr. No.	Districts→ Years ↓	Aurangabad	Parbhani	Osmanabad	Beed	Nanded
1	1976	201	461	435	254	612
2	1977	663	964	1985	564	1204
3	1978	1479	1862	2492	1602	3244
4	1979	1230	1337	3798	1556	2197
5	1980	1286	2402	4249	1348	3157
6	1981	1185	1344	2365	1189	1828
7	1982	1595	2080	3418	978	1876
8	1983	1288	2074	3351	942	3708
9	1984	1531	1288	2297	854	861
10	1985	909	1651	3410	1048	2938
11	1986	751	1393	2149	674	2111
12	1987	1385	1037	668	516	1329
13	1988	1652	1880	3876	1396	3389

14	1989	814	944	1199	471	676
15	1990	1763	2151	3295	1086	3875
16	1991	1438	2424	3443	1128	2844
17	1992	784	1610	1403	383	1588
18	1993	1904	4225	5656	1674	4846
19	1994	1481	2728	3682	1127	3694
20	1995	1078	2583	2641	713	3431
21	1996	919	2737	4023	1331	2046
22	1997	1550	331	4267	1155	3769
23	1998	716	2555	2437	557	2195
24	1999	858	2025	3487	580	2344
25	2000	754	2555	3244	889	2655
26	2001	530	2386	3383	1047	2566
27	2002	578	2318	2572	434	2107

Table-4.1A: Elementary Statistics of Production of Kharif Jawar Time Series Data of Marathwada Region for 27 Years (1976-2002)

Cities:	**Aurangabad**	**Parbhani**	**Osmanabad**	**Beed**	**Nanded**
Mean:	1043.2	1926.7	2903.4	1092.4	2359.9
S.D.:	263.5	142.1	287.3	234.9	224.9
C.V.:	25.2	7.3	9.9	21.5	9.5

Table-4.1B: Linear Regression Analysis of Production of Kharif Jawar Time Series Data To Determine Trend Eq(7)

District	**Coefficients**		**Standard Error**	**t Stat**	**Significance**
Aurangabad	β_0	1319.95	87.97	15.00	S
	β_1	-19.76	5.49	-3.60*	S
Parbhani	β_0	1899.47	58.16	32.66	S
	β_1	1.95	3.63	0.54	NS
Osmanabad	β_0	3160.73	102.48	30.84	S
	β_1	-18.38	6.40	-2.87*	S
Beed	β_0	1461.38	47.06	31.05	S
	β_1	-26.35	2.94	-8.97*	S
Nanded	β_0	2671.17	59.09	45.20	S
	β_1	-22.23	3.69	-6.03*	S

*t =2.060 is the critical value for 25 d f at 5% L. S. * shows the significant value*

A look at the table 4.1A 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 Parbhani districts but ***significant*** in rest of the four districts. A simple look at the mean values shows that a classification as

C1 = {Aurangabad, Beed}

C2 = {Nanded, Parbhani, Osmanabad}

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 4.2A). Such an analysis requires an assumption of AR (Auto-regressive) model [12] 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

$X_t = C + \Phi X_{t-h} + €_t$, h = 0, 1, 2, … .20 … (8)

Table-4.2A : Auto Variances: Individual Column Treated As Ordinary Time Series For Lag Values (H = 0 , 1 , 2 ,………20) About Production of Kharif Jawar Time Series Data

lag h	**Aurangabad**	**Parbhani**	**Osmanabad**	**Beed**	**Nanded**
0	180868.6	654369.3	1317106.6	151637.8	1106913.4
1	28567.6	73563.6	-133606.8	-872.7	-259453.4
2	-4685.0	195348.3	42984.7	-26026.6	145561.0
3	49746.4	130784.7	425931.3	31289.3	300887.1
4	3925.1	-247195.7	-490092.3	-45099.7	-306131.8
5	-45731.3	172232.0	164936.0	-10811.0	405623.7
6	17905.0	-104628.6	-67913.9	-15435.6	-420573.5
7	-31992.5	-26939.2	-582772.8	-24541.7	102403.2
8	-80763.1	220357.3	208929.4	-8674.7	74816.6
9	37713.8	-67591.8	-219308.3	-25405.4	-388887.9
10	26422.0	224638.9	-99902.3	29113.2	363147.6
11	-56479.3	172750.8	651313.3	-10505.6	-84662.3

12	-25413.3	24909.2	-249977.3	9213.2	149494.2
13	20620.0	293424.0	439021.2	52835.9	126377.8
14	-88276.4	-90860.2	335988.1	-24199.8	-62475.9
15	28908.6	12218.6	103715.3	75882.8	342553.4
16	-1204.0	37994.1	234421.2	42343.2	-59422.6
17	-117771.3	-410954.1	-353543.5	-57853.8	-253309.0
18	-94375.3	40647.6	-63916.2	22428.3	-297127.8
19	-25794.8	-181328.1	-32201.1	-18564.9	-69114.4
20	-85395.0	136712.7	-367765.4	-108767.3	-82353.9

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

Corr. Coefficient	**-0.853***	-0.179	**-0.711***	**-0.969***	**-0.888***

*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** was not found ***significant*** in Parbhani district only showing that the time series can be reasonably assumed to be **stationary.** The coefficient is significant, with negative value showing that Aurngabad, Osmanabad Beed and Nanded has been experiencing significantly declining production 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	45778.37	7792.56	5.87	**S**
	β_1	-4756.24	666.57	**-7.14***	**S**
Parbhani	β_0	2770.69	3632.30	0.76	NS
	β_1	-246.01	310.71	-0.79	NS
Osmanabad	β_0	38846.36	9110.32	4.26	**S**
	β_1	-3432.83	779.30	**-4.41***	**S**
Beed	β_0	45038.88	1896.55	23.75	**S**
	β_1	-2769.97	162.23	**-17.07***	**S**
Nanded	β_0	34926.17	3497.47	9.99	**S**
	β_1	-2513.21	299.17	**-8.40***	**S**

*t =2.093 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. Dr. 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) (1991), 324-326.*
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) (2006), 1-10.*
6. *Bable, B. L. and Pawar D., D.. Covariance Stationary Time Series; Multidisciplinary Bi-lingual International Journal 'Innovation', 2(7) (2017), 28-32.*
7. *Bable, B. L. and Pawar D., D.. Properties of Weak and Strict Stationary Time Series;. Multidisciplinary International Research Journal Asian Resonance, 4(3) (2015), 54-58.*
8. *Bable, B. L. and Pawar D., D.. Trends in Scalar Time Series;. Multidisciplinary International Research Journal Periodic Research, 4(1) (2015), 41-46.*
9. *Bable, B. L. and Pawar D., D.. Formation of Trends in Scalar Time Series;. International Research Journal Agricultural Economics and Statistics, 5(1) (2014), 111-115.*
10. *Bhattacharya, G. K. and Johanson R. A., Statistical Concepts and Methods. John Wiley and Sons, New York (1977).*
11. *Das, P. K., Statistics and Its Applications to Meteorology. Indian J. pure appl. Math., 26(6) (1995) , 531-546.*
12. *Fuller W. A., Introduction to Statistical Time Series. John Wiley and Sons, New York (1976).*
13. *Hooda R. P., Statistics for Business and Economics. Machmillan India Ltd., New Delhi (2003).*
14. *Patil, V. D. and Kalyanker S. P., Impact of Research on Productivity of Agricultural in Maharastra. Agri. Situ. India, 47(5) (1982), 321-326.*
15. *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) (1993), 425-426.*