P: ISSN NO.: 2394-0344 RNI No.UPBIL/2016/67980 VOL-I* ISSUE-IX* December- 2016
E: ISSN NO.: 2455-0817 Remarking An Analisation

Station arty in Time Series

Abstract

In this paper, a time series {X(t, ω), t Є T} on (Ω , C, P) is explained. Where X is a random variable (r. v.).The properties of stationary time series with supporting real life examples have been taken.Production of wheat series for 33 years from five districts of Marathwada region in Maharastra State were analyzed.

A preliminary discussion of properties of time series precedes the actual application to production of wheat data.

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

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

D. D. Pawar
Associate Professor & Chairman,
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 and conclusions have been drawn by testing methodology of hypothesis. In this article we have used production of wheat data of 1970 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.
2. To develop algorithms for analyzing time series, which use the characterizing theorems.
3. To interpret the results of characterizations, in real economic and social terms.

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 2.2

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 satisfy.

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

Where, the equality must hold for all possible sets of indices t_i and (t_i + 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.

Theorem 2.1

If {X_t: t Є T}, is strictly stationary with $E\{|X_t\} < \alpha$ and

$$\left.\begin{array}{l} E\{|X_t - \mu|\} < \beta \text{ then ,} \\ E(X_t) = E(X_{t+h}), \text{ for all t, h} \end{array}\right\} \ldots\ldots (2)$$

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

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 2.4

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 2.5

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 $\Upsilon(h)$.

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

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

Definition 2.6

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

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

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

where μ is mean.

Remark 2.1

For a stationary time series the variance at time (t + h) is same as that at time t. 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 of wheat 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. That is $\Upsilon(h) = \Upsilon(-h)$.

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 $t_0 = t_1 - h$,

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

proved.

Theorem 2.3

Show that, with usual notation $|\Upsilon(h)| \le \Upsilon(0)$.

Proof

Since $\rho(h) = \Upsilon(h)/\Upsilon(0)$ and since $\Upsilon(0) = 1/n \, \Sigma(X_i - \bar{X})^2 \ge 0$, we wish to show that $|\Upsilon(h)| \le \Upsilon(0)$.

Consider first two random samples $X_1, X_2 \ldots X_n$ and $Y_1, Y_2, \ldots Y_n$ and the quantity

$$\sum_{i=1}^{n} [a(X_i - \bar{X}) + (Y_i - \bar{Y})]^2 \quad \ldots (7)$$

$$= a^2\sum_{i=1}^{n}(X_i - \bar{X})^2 + \sum_{i=1}^{n}(Y_i - \bar{Y})^2 + 2a\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) \quad \ldots (8)$$

Taking the derivative of (8) w r t. a and setting it equal to zero gives us

$$2a\Sigma(X_i - \bar{X})^2 + 2\Sigma(X_i - \bar{X})(Y_i - \bar{Y}) = 0 \quad \ldots (9)$$

$$\Rightarrow a = \frac{-\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{\Sigma(X_i - \bar{X})^2}$$

Since the second derivative with respective to a is $\Sigma(X_i - \bar{X})^2 > 0$ (assuming a non-trivial sample of one repeated constant value), we see that this value of a in fact minimizes (7).

Since (8) = (7) and (7) ≥ 0, putting the value of a in eqn. (8) gives us

$$\frac{-[\Sigma(X_i - \bar{X})(Y_i - \bar{Y})]^2}{\Sigma(X_i - \bar{X})^2} + \Sigma(Y_i - \bar{Y})^2 \ge 0 \quad \ldots (10)$$

(13) implies that

$$[\Sigma(X_i - \bar{X})(Y_i - \bar{Y})]^2 \le [\Sigma(X_i - \bar{X})^2]\,[\Sigma(Y_i - \bar{Y})^2]$$

This is known as the cauchy – Schwartz inequality.

Taking the square root of both sides, we have

$$|\Sigma(Xi-X)(Yi-\bar{Y})| \le \{[\Sigma(Xi-\bar{X})2]\,[\Sigma(Yi-Y)2]\}1/2 \ldots (11)$$

In (14), replace $\bar{Y}$ with $\bar{X}$, replace n with n-h, and replace Y_i with X_{i+h}

$$\left|\sum_{i=1}^{n-h}(X_i - X)(X_{i+h} - X)\right| \le \left\{\sum_{i=1}^{n-h}(X_i - \bar{X})^2 \sum_{i=1}^{n-h}(X_{i+h} - \bar{X})^2\right\}^{1/2}$$

Note that the resulting right-hand side will be made larger if we sum those squares over all indices from 1 to n. Thus,

$$\left|\sum_{i=1}^{n-h}(X_i - X)(X_{i+h} - X)\right| \le \left\{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(X_i - X)^2\right\}^{1/2} = \sum_{i=1}^{n}(X_i - X)^2$$

Multiplying both sides by 1/n, we have $|\Upsilon(h)| \le \Upsilon(0)$.

Theorem 2.4

Let X_t's be independently and identically distributed with $E(X_t) = \mu$ and $var(X_t) = \sigma^2$

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

This process is stationary in the strict sense.

Testing Procedure

3.1: Inference concerning slope (β_1)

We set up null hypothesis for test statistic 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.

$$t_{n-2} = (b - \beta)\{\sqrt{S_{tt}}\}/s_e,$$

Where, $s_e^2 = \dfrac{S_{xx} - (S_{xt})^2 / S_{tt}}{n-2}$

Example of Time Series

Production of wheat data of five districts namely Aurangabad, Parbhani, Osmanabad, Beed and Nanded in Marathwada region were collected. The data were collected from Epitome of Agriculture (Part-II) for Maharastra Quarterly Bulletin of Economics and Statistics, Bombay [2]. Hence we have five dimensional time series t_i, i = 1, 2, 3, 4, 5 districts respectively.

Over the years many scientists have analyzed rainfall, temperature, humidity, agricultural area, production and productivity of region of Maharastra state, [1, 3, 4, 5, 7, 10]. 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].

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 no any scatter of values around the mean indicating that all the series are 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 stationary in Osmanadad, Beed and Nanded districts. Hence X_t is not having trend in Osmanadad, Beed and Nanded districts.

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

where,

1. X_t is the annual production of wheat 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 of wheat 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 of wheat values for different districts were input as a matrix to the software. Defining

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

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

$\Upsilon(h) = \mathrm{cov}(A, B)$ 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 $\Upsilon(h)$ and 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 for five the district. Table-4.2C was obtained by regressing values of $\Upsilon(h)$ and h, using “Data Analysis Tools” provided in MS Excel. Table 4.2A formed the input for table 4.2C. In other wards, $\Upsilon(h)$ are all non zero in five districts, in the production of wheat series in five districts trends were found showing that X_t,

X_{t+h} are independent in production of wheat series of these districts and there is no trend in that series. Hence in these districts production of wheat series X_t ‘s were stationary. These series are having no trend.

Conclusion

The number of production of wheat series showed to be not significant (coefficients) intercepts and slope. The districts Aurangabad, Parbhani, Osmanabad, Beed and Nanded individually showed stationary in nature of these series.

Generally it is expected, production of wheat (annual) over a long period at any region to be not stationary time series. The result does not conform with these series, of five districts.These series have not found trend. The regression coefficient (β_1) is not significant.

Analysis: Production of wheat Data

The strategy of analyzing individual time series as scalar series has been adapted here for production of wheat.

Table 4.1 contains the results for scalar series approach.

The model considered was:

$X_i(t) = (\beta_0)_i + (\beta_1)_i t + \in_i(t), \quad i = 1, 2 \dots 5 \quad \dots (12)$

Where X_i is the annual production of wheat series, t is the time variable, β_0 = the intercept, β_1 = the slope, $€_i$ is the random error. production of wheat data X is the dependent variable and time t in years is the independent variable.

Table-4.1

Districtwise Production of Wheat Data in (Hundred Tonnes)

Sr. No.	*Year*	Aurangabad	Parbhani	Osmanabad	Beed	Nanded
1	1970	277	202	210	182	118
2	1971	191	138	156	116	56
3	1972	347	282	248	241	103
4	1973	71	87	67	43	17
5	1974	487	336	404	237	217
6	1975	759	489	483	282	269
7	1976	1190	662	793	407	382
8	1977	857	478	517	492	363
9	1978	857	478	517	492	363
10	1979	881	566	596	389	319

11	1980	618	538	574	518	351
12	1981	1166	540	379	448	235
13	1982	822	490	473	396	288
14	1983	651	469	406	267	287
15	1984	1027	600	670	327	316
16	1985	504	533	389	267	270
17	1986	344	346	268	178	173
18	1987	304	242	122	96	82
19	1988	494	337	253	200	153
20	1989	869	519	543	309	283
21	1990	829	589	524	300	269
22	1991	732	450	481	324	215
23	1992	450	382	128	95	153
24	1993	821	428	259	168	279
25	1994	941	635	582	280	292
25	1995	839	689	318	238	248
27	1996	650	623	433	387	239
28	1997	794	889	672	395	294
29	1998	436	314	177	204	92
30	1999	1149	1029	752	463	359
31	2000	1087	1106	554	475	391
32	2001	894	960	602	471	286
33	2002	765	1133	497	371	347

Table-4.1A

Districtwise Production of Wheat Data in (Hundered Tonnes) of Marathwada Region for 33 Years (1970-2002). Simple Descriptive Statistics

Cities:	Aurangabad	Parbhani	Osmanabad	Beed	Nanded
Mean:	700.09	532.09	425.67	304.79	245.73
S.D.:	82327.23	64813.11	35127.01	16473.68	9856.56
C.V.:	755.07	669.96	493.22	337.76	261.26

Table-4.1B

Linear regression analysis of production of wheat data to determine trend Eq(12).

Districts	Coefficients		Standard Error	t Stat	Signi. or Not
Aurangabad	β_0	510.16	97.94	5.21	S
	β_1	11.17	5.03	**2.22**	**S**
Parbhani	β_0	217.97	67.63	3.22	S
	β_1	18.48	3.47	**5.32**	**S**
Osmanabad	β_0	346.44	66.93	5.18	S
	β_1	4.66	3.43	**1.36**	**NS**
Beed	β_0	257.28	46.15	5.58	S
	β_1	2.79	2.37	**1.18**	**NS**
Nanded	β_0	193.66	34.88	5.55	S
	β_1	3.06	1.79	**1.71**	**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-4.1A shows that all of them have not similar values of CV.Which indicates that their dispersion is not almost identical. Trends were found to be significant in two districts Aurangabad and Parbhani. A simple look at the mean values shows that a classification as

C1 = {Aurangabad, Parbhani}

C2 = {Osmanabad, Beed, Nanded} could be quite feasible.

There is no linear trend, in three districts which reasonably taken as evidence of series being stationary in these districts Osmanabad, Beed, Nanded individually.

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 [9] Eq (13). 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 … (13)

Table-4.2 A

Auto variances: Individual column treated as ordinary time series for lag values (h = 0 , 1 , 2 , … 20) about production of wheat data

lag h	Aurangabad	Parbhani	Osmanabad	Beed	Nanded
0	82327.2	64813.1	35127.0	16473.7	9856.6
1	38442.8	38063.1	8922.4	10029.0	5277.6

2	17897.0	30810.0	5452.2	6844.0	2867.1
3	2701.2	21689.1	-1303.6	3787.3	705.6
4	-10632.8	10840.8	-4026.4	412.6	-1084.8
5	-1618.2	20587.5	1682.4	-1223.1	-1412.8
6	-6550.7	10752.6	-1260.6	-3725.4	-1333.9
7	-16802.6	5191.3	-8507.8	-8525.7	-3046.6
8	-23188.5	-1640.3	-6630.5	-8064.3	-3066.6
9	-18370.0	-2212.9	-10399.3	-8483.3	-3048.1
10	-13013.6	1636.3	-2306.1	-6868.3	-2357.2
11	-24252.7	1215.7	-8734.1	-5527.2	-1855.5
12	-1608.0	665.0	-8056.9	-4605.3	-2897.2
13	9093.1	-674.6	-3666.3	-3685.7	-1672.3
14	6369.4	1363.7	-2151.5	-3076.3	106.7
15	29496.2	11650.0	10967.2	-1427.5	1943.5
16	19793.1	17092.3	-2259.6	521.4	2426.0
17	12631.2	19097.3	5205.9	2568.7	3690.9
18	30645.3	23867.3	1367.3	2840.5	1252.4
19	18114.1	21676.6	3841.9	10568.3	2020.6
20	3996.6	27919.3	11017.0	10055.9	1824.3

Table-4.2B

Correlation coefficient between h and Auto covariance is

Districts	Aurangabad	Parbhani	Osmanabad	Beed	Nanded
Corr. Coeff.	**0.240**	**-0.092**	**0.204**	**0.136**	**0.162**

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

A look at the table-4.2B shows that correlation's between Υ_{ij} **(h)** and **h** were found to be not ***significant*** in five districts showing that the time series can be reasonably assumed to be **stationary** i.e. these districts have no trend**.**

Table-4.2C

Linear regression analysis of lag values vs covarianc.

		Coefficients	**Standard Error**	**t Stat**	
Aurangabad	β_0	-4185.58	8531.16	-0.49	NS
	β_1	746.93	712.17	**1.05**	**NS**
Parbhani	β_0	14969.90	5778.44	2.59	S
	β_1	-189.56	482.38	**-0.39**	**NS**
Osmanabad	β_0	-2903.72	3041.64	-0.95	NS
	β_1	224.90	253.91	**0.89**	**NS**
Beed	β_0	-1883.46	2942.52	-0.64	NS
	β_1	143.26	245.64	**0.58**	**NS**
Nanded	β_0	-709.36	1193.29	-0.59	NS
	β_1	69.18	99.61	**0.69**	**NS**

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

Significant trends were not found in five districts.

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, Epitome of Agriculture (Part-II) for Maharstra State and Maharastra Quarterly Bulletin of Economics and Statistics, Bombay (1970-2002).
3. 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.
4. Bable, B. L. and Pawar D. D., Trends of District-wise Scalar Time Series; International Journal of Statisticsa and Mathematica, ISSN:, Valume 1, Issue 1, (2011) pp 01-07.
5. Bable, B. L. and Pawar D. D., Vector Time Series: The Concept and Properties to the Vector Stationary Time Series; Internat. Res. J. of Agric. Eco. and Stat., 3(1) (2012), 84-95.
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) (1995), 531-546.
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. Waghmare, P. R. and Katapallewar B. N., District-wise 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.