Covariance Stationary Time Series

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

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

Abstract

In this paper, a time series $\{X_t(\omega), t \in T\}$ on (Ω, C, P) is explained. Where X is a random variable (r. v.). The properties of covariance and strict stationary time series with supporting real life examples have been taken and conclusions have been drawn by using methodology of testing hypothesis. Number of rainy days and rainfall data for 25 years from Osmanabad District of Maharastra State were analyzed.

A preliminary discussion of properties of time series and possible testing methodology for stationary property precedes the actual application to number of rainy days and rainfall data.

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

Introduction

Our aim here is to illustrate a few properties of weak and strict 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 rainfall data and number of rainy days of 1981 to 2004 at Beed district to illustrate most of properties theoretically established.

Objective of the Study

1. To develop theory of time series,
2. To develop algorithms for analyzing time series, and
3. To interpret the results of characterization in real social terms.

Basic Concepts

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

Definition 2.1

Probability space: A probability space is a triplet (Ω, C, P) where

1. Ω is a set of all possible results of an experiment;

2. C is class of subsets of Ω (called events) forming a σ- algebra, i.e.
   i). $\Omega \in C$,
   ii). $A \in C \Rightarrow A^c \in C$,
   iii). $\bigcup_{j=1}^{\infty} Aj \in C$, for any sequence $\{A1, A2, \ldots\} \subseteq C$;

3. P: $C \rightarrow [0, 1]$ is a function which assigns to each event $A \in C$ a number $P(A) \in [0, 1]$, called the probability of A and such that
   i). $P(\Omega) = 1$,
   ii) If $\{Aj\}_{j=1}^{\infty}$ is a sequence of disjoint events, then $P(\bigcup_{j=1}^{\infty} Aj) = \Sigma_{j=1}^{\infty} P(Aj)$

Definition 2.2 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 $t \in T$, 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 $\{X_t : t \in T\}$, of random variables[7].

Definition 2.3 Stationary Time Series

The plot of a time series over a time interval [t, t+h] may sometimes closely resemble a plot at another interval [s, s+h]. This implies that there is temporal homogeneity in the behavior of the series, which is called stationary. For example, the number of personal bankruptcies may be stationary in monthly data. This means that the time series between January to March in one year may resemble June to August of another year. A stationary series should have no discernible trends. For precise definition of stationary, some concepts from probability theory, which are developed later, are needed. An imprecise operational definition of stationary time series is as follows: When the mean: E(Xt), the variance: Var(Xt) and all autocavariances of specified lags(say h): Cov(Xt, Xt+h) do not depend on t, the time at which they are measured, we have a 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 Weak Stationary or Covariance Stationary Time Series

A weaker concept of stationary allows the joint distributions to change some what over time, but requires that E(Xt), Var(Xt) do not change. Also, Cov(Xt, Xt+h) is required to be a function of lag length only, not depend on time t at which it is measured. Many stationary series are also called covariance stationary, wide-sense stationary, or second order stationary in the literature. For the multivariate normal joint distribution, weak and strict stationarity are equivalent.

For a random process to be weakly stationary – also called covariance stationary – there are three requirments:

1. $E(x_t) = \mu < \infty$ for all t, & is not function of t.
2. $E(x_t-\mu)^2 < \infty$ for all t, & is not function of t. } … (1)
3. $E(x_t - \mu)(x_{t-h} - \mu) = \Upsilon_h < \infty$ for all t & is not function of t.

Definition 2.5 Strictly Stationary Time Series

Strict stationary time series is a stronger concept where the properties are unaffected by a change of time origin. It requires that the joint distribution function F,

$F[\, x(t_1), x(t_2), \dots x(t_n)] = F[\, x(t_{1+h}), x(t_{2+h}), \dots x(t_{n+h})]$ … (2)

For all choices of time points x_1 to x_n and for all h. If the first two moments of strictly stationary process exist, that is they are finite it is also covariance stationary and it satisfies the three conditions (1). Stationary implies stable relations and the object of any theory is to obtain stable relationship among variables.

Main Results

Theorem 3.1

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 } … (3)

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

Proof

Proof follows from definition (2.5).

In usual cases above equation (3) is used to determine that a time series is stationary i.e. there is no trend.

Definition 3.1: 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 $\Upsilon(h)$.

$\Upsilon(h) = Cov(X_t, X_{t+h}) = E\{X_t - \mu\}\{X_{t+h} - \mu\}$ … (4)

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

Theorem 3.2

(Properties of the autocovariance function). Let $\{X_t: t \in T\}$ be a second- order stationary process. The autocovariance function $\Upsilon(h)$ of the process satisfies the following properties:

1. $\Upsilon(0) = Var(X_t) \geq 0$, for all $t \in T$;
2. $\Upsilon(h) = \Upsilon(-h)$, for all $h \in Z$ (i.e. $\Upsilon(h)$ is an even function of h);
3. $\Upsilon(h) \leq \Upsilon(0)$, for all $h \in Z$;
4. The function $\Upsilon(h)$ is positive semi-definite, i.e.

$$\sum_{j=1}^{n}\sum_{k=1}^{n} a_k \Upsilon(t_j - t_k)\, a_j \geq 0,$$ for any positive integer n and for all the vectors

$a = (a_1, a_2, \dots a_n)$ and any set of indices $(t_1, t_2 \dots t_n) \in T$ such that $(t_j - t_k) \in X$.

Definition 3.2: 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

$$P(h) = \frac{E\{X_t - \mu\}\{X_{t+h} - \mu\}}{[E\{X_t - \mu\}^2 E\{X_{t+h} - \mu\}^2]^{1/2}} \quad \dots (5)$$

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

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]. Thus, the auto correlation at lag h is

$$\rho(h) = \frac{\Upsilon(h)}{\Upsilon(0)} \quad \dots (6)$$

Remark 3.2

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

For application, attempts have been made to establish that stationary satisfy equation (2) and (6).

Theorem 3.3

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 $t_0 = t_1 - h$,

$\Upsilon(h) = E\{X_{t0} X_{t0+h}\} = E\{X_{t1} X_{t1-h}\} = \Upsilon(-h)$ … (7)

proved.

Definition 3.3: Positive Semi-Definite

A function f(x) defined for $x \in X$ is said to be positive semi-definite if it satisfies

$$\sum_{J=i}^{n}\sum_{k=1}^{n} a_k^T f(t_j - t_k)\, a_j \geq 0,$$

for any set of real vectors $(a_1, a_2, \dots a_n)$ and any set of indices $(t_1, t_2 \dots t_n) \in T$ such that $(t_j - t_k) \in X$.

Theorem 3.4

The covariance function of stationary time series $\{X_t: t \in T\}$ is positive semi-definite function in that

$$\sum_{j=1}^{n}\sum_{k=1}^{n} a_k \Upsilon(t_j - t_k)\, a_j \geq 0,$$

For any set of real ($a_1, a_2, \ldots a_n$) and any set of indices ($t_1, t_2 \ldots t_n$) $\in$ T.

Proof

The result can be obtained by evaluating the variance of

$$X = \sum_{j=1}^{n} a_j X_{t_j}.$$

For this without loss of generality $E(X_t) = 0$. It shows that the variance of a random variable is non-negative i.e. $V(X) \geq 0$.

$$V(X) = V(\Sigma\ a_j X_{t_j}) \geq 0$$

$$= E\left(\sum_{j=1}^{n} a_j X_{t_j}\right)\left(\sum_{k=1}^{n} a_j X_{t_j}\right) \geq 0,$$

$$= \sum_{J=1}^{n}\sum_{k=1}^{n} a_j a_k E\{X_{t_j} X_{t_k}\} \geq 0,$$

$$= \sum_{j=1}^{n}\sum_{k=1}^{n} a_k \Upsilon(t_j - t_k)\, a_j \geq 0 \quad \ldots (8)$$

Hence proved.

Theorem 3.5

$|\rho_{12}(h)| \leq 1.$

Proof

If we set n = 2, in the equation (8) to obtain,

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_j \Upsilon(t_i - t_j)\, a_i = a_1^2 \Upsilon_{11}(0) + a_2^2 \Upsilon_{22}(0) + 2a_1 a_2 \Upsilon(t_1 - t_2) \geq 0.$$

$a_1^2\Upsilon_{11}(0) + a_2^2\Upsilon_{22}(0) \geq -2a_1 a_2 \Upsilon_{12}(t_1 - t_2)$, since $\Upsilon_{11}(0) = \Upsilon_{22}(0)$

$$1/2(a_1^2 + a_2^2) \geq \frac{-a_1 a_2 \Gamma_{12}(t_1 - t_2)}{\Gamma_{11}(0)}$$

Now, let $a_1 = a_2 = 1$ and $t_1 - t_2 = h$,

$$1 \geq \frac{-\Upsilon_{12(h)}}{\Upsilon_{11}^{(0)}} = -\rho_{12}(h) \quad \ldots (9)$$

Similarly, $-a_1 = a_2 = 1$; it shows that

$$P_{12}(h) \leq 1 \quad \ldots (10)$$

From (9) and (10) we get

$$|\rho_{12}(h)| \leq 1.$$

Hence proved.

Theorem 3.6

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

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

This process is stationary in the strict sense.

Testing Procedure

Inference Concerning Slope (β_1)

We set up null hypothesis for testing $H_0: \beta_1 = 0$ 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. The hypothesis H_0 is not significant for both the values of t for 22 and 14 d. f. for the district.

$$t_{n-2} = \beta_1 / Syx\{1/\Sigma(X - \bar{X})\}^2$$

where β_1 is the slope of the regression line.

Example of Time Series

Rainfall and number of rainy days data were collected from Vasantrao Naik Marathwada Agricultural University, Parbhani[1]. Hence we have two dimensional time series t_i, i = 1, 2 corresponding to the district Osmanabad. 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.

Over the years many scientists have analyzed rainfall, temperature, humidity, agricultural area, production and productivity of region,[2, 3, 4, 6]. 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.[8] and for testing correlation coefficient Bhattacharya G.K. and Johanson R.A. [5]**.**

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 5.1 C and Table 5.2 C. Elementary statistical analysis is reported in Table 5.1 A and Table 5.1 B. 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.

Table5.1C shows that the model,

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

When applied to the data indicates $H_0: \beta_1 = 0$ is true. Hence X_t is a not having trend for two series of the district. where,

1. X_t are the annual rainfall or number of rainy days series.
2. t is the time (years) variable.
3. ϵ is a random error term normally distributed with mean 0 and variance σ^2.

Rainfall or number of rainy days 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 Table-5.2A. Rainfall or number of rainy days values for district was input as a matrix to the software. Defining

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

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

$\Upsilon(h) = cov(A, B)$ were computed for various values of h. Since the time series constituted of 24 values, at least 9 values were included in the computation.The relation between $\Upsilon(h)$ were examined using model, Table-5.2C.

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

The testing shows that, both the hypothesis $\beta_0 = 0$ and $\beta_1 = 0$ test is not positive. Table-5.2C was obtained by regression 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 rainfall and number of rainy days series of the district, trends were not found showing that X_t, X_{t+h} are not dependent in both the series of the district and there is no trend in that series.

Conclusion

It was observed that t values are therefore not significant for both the series of the district, i.e. concluded that X_t does not depend on t for both the series of the district [5]. Similarly, Υ_{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 not positive for t and h for both the series of the district.

Generally it is expected, rainfall or number of rainy days (annual) over a long period at any region to be stationary time series. The results are confIrm with these series in the district i, e. in the district trends were not found in both the series.

Rainfall or Number of Rainy Days Time Series Treated As Scalar Time Series

Table 5.1 contains the results for scalar series approach.

The model considered was:

$$X_i(t) = (\beta_0)_i + (\beta_1)_i t + \epsilon_i(t), \quad i = 1, 2. \quad \ldots (11)$$

Where X_i is the annual rainfall or number of rainy days 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-5.1: Annual Rainfall and Number of Rainy Days Data of Osmanabad District.

S. No.	District→ Years ↓	Number of Rainy Day	Rainfall
1	1981	51	812.8
2	1982	35	567.5
3	1983	51	1013.5
4	1884	25	363.7
5	1985	33	563
6	1986	33	538
7	1987	45	582.2
8	1988	66	1263.6
9	1989	38	1041.4
10	1990	45	916.4
11	1991	30	467.1
12	1992	28	485.7
13	1993	45	708.6
14	1994	33	423.6
15	1985	30	688.4
16	1996	40	717.8
17	1997	39	575.7
18	1998	59	1314
19	1999	35	591.4
20	2000	44	959.6
21	2001	41	594.5
22	2002	39	633.7
23	2003	29	428
24	2004	44	752

Decriptive Statistics

Table-5.1 A

Elementary Statistics of Observed Minimum, Maximum, Mean, Standard Deviation (S.D.) and Coefficient of Variation (C.V.) Of Number of Rainy Days of Osmanabad District.

Nanded District	Minimum (mm)	Maximum (mm)	Mean (mm)	S.D. (mm)	C.V. %
Rainydays	363.7	1314	708.4	251.7	35.5

Table-5.1 B

Elementary Statistics of Observed Minimum, Maximum, Mean, Standard Deviation (S.D.) and Coefficient of Variation (C.V.) of Annual Rainfall of Osmanabad District.

Nanded	Minimum (mm)	Maximum (mm)	Mean (mm)	S.D. (mm)	C.V .
Rainfall	25	66	39.9	9.7	24.4

Table-5.1 C

Linear Regression Analysis of Data To Determine Trend.

Nanded District	Coefficients		Standard Error	t Stat	Significance	P Value	Lower 95%	Upper 95%
Rainydays	β_0	40.9	4.3	9.5	S	0.0	32.0	49.8
	β_1	-0.1	0.3	-0.3	NS	0.8	-0.7	0.5
Rainfall	β_0	721.6	110.7	6.5	S	0.0	492.0	951.3
	β_1	-1.1	7.7	-0.1	NS	0.9	-17.1	15.0

*t =2.04 is the critical value for 20 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 both the series of the district. In absence of linear trend, with reasonably low CV values can be taken as evidence of series being stationary series individually in 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[7] Eq(12). 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

$X_t = C + \Phi X_{t-h} + \epsilon_t$, h = 0, 1, 2, … .15 … (12)

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.

Beed District		
lag h	Number of rainy days	Rainfall
0	95.0	63367.1
1	-9.1	-2907.2
2	6.4	10220.9
3	-30.6	-22048.9
4	-37.1	-30517.1
5	8.3	-6195.8
6	-19.6	-14828.1
7	8.4	6203.2
8	12.0	9110.5
9	-7.1	11601.0
10	43.9	35471.6
11	4.0	-6155.6
12	-1.0	8891.9
13	-13.6	-21054.5
14	-37.8	-39973.7
15	-7.1	9785.6

Table-5.2 B

Correlation Coefficient between H and Auto Covariance is

	Number of rainy days	Rainfall
Corr. Coefficient	-0.29	-0.25

Correlation coefficient r = 0.433 is the critical value for 14 d f at 5% L. S. *Shows the significant value.

Correlation's between Υ_{ij} (h) and h were found tb be not *significant* in both the series of the district. It is seen that the time series can be reasonably assumed to be stationary i.e. not having trend**.**

Table-5.2 C

Linear Regression Analysis of Lag Values H Vs Covariance

Nanded District	Coefficients		Standard Error	t Stat	Significance	P Value	Lower 95%	Upper 95%
Rainydays	β_0	15.8	15.3	1.0	NS	0.3	-17.1	48.7
	β_1	-2.0	1.7	-1.1	NS	0.3	-5.7	1.7
Rainfall	β_0	10502.8	12093.1	0.9	NS	0.4	-15434.3	36439.9
	β_1	-1309.0	1373.7	-1.0	NS	0.4	-4255.2	1637.3

*t =2.1 is the critical value for 14 d f at 5% L. S. * shows the significant value*

Conclusion

It was observed that t values are therefore not significant for both the series of the district, i.e. concluded that X_t does not depend on t for both the series of the district [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 not positive for t and h for both the series of the district.

Generally it is expected, rainfall or number of rainy days (annual) over a long period at any region to be stationary time series. The results are conform with these series in the district i, e. in the district trends were not found in both the series.

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. of Agricultural Economics and Statistics, ISSN:, 3(1) (2014), 111-115.*
5. *Bhattacharya, G. K. and Johanson R. A., Statistical Concepts and Methods. John Wiley and Sons, New York (1977).*
6. *Das, P. K., Statistics and Its Applications to Meteorology. Indian J. pure appl. Math., 26(6) (1995), 531-546.*
7. *Fuller W. A., Introduction to Statistical Time Series. John Wiley and Sons, New York (1976).*
8. *Hooda R. P., Statistics for Business and Economics. Machmillan India Ltd., New Delhi (2003).*