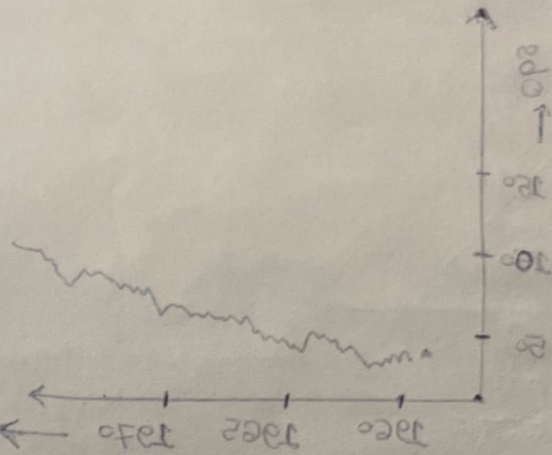


(ii) $\frac{b}{a} > 1$ (Exponential growth)

* Time	observation
1960	x_1
1965	x_2
1975	x_3
1980	x_4
⋮	⋮



Exponential growth
↓
Exponential decay

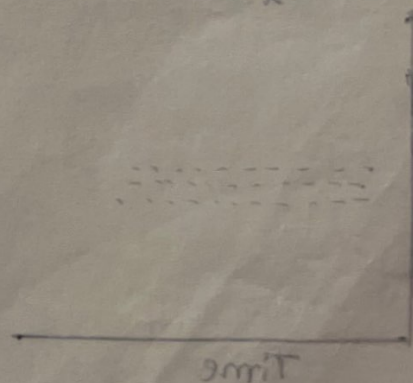
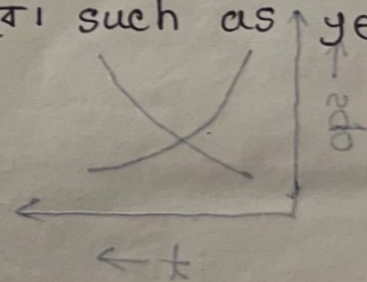
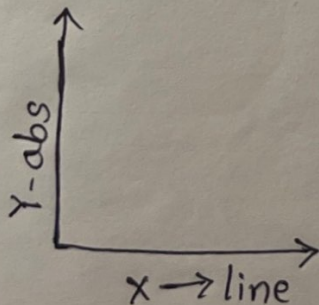
* একটি time series data হবে না।

* Time series data তে time এর equal interval থাকতে হবে।

□ Time series plot এ x-axis এ time y-axis এ observation থাকবে।

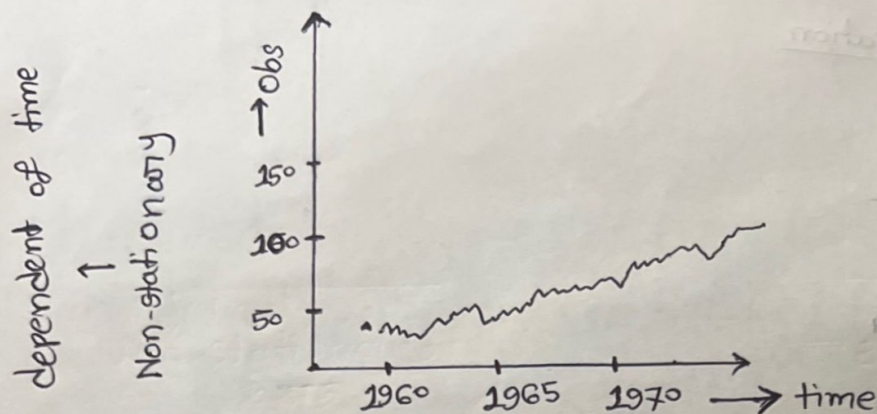
Characteristic of Time Series Data:

(1) Time interval same (equal) থাকবে। such as yearly, monthly, quarterly etc



Exponential growth
↓
Exponential decay

(ii) Upward/downward trend

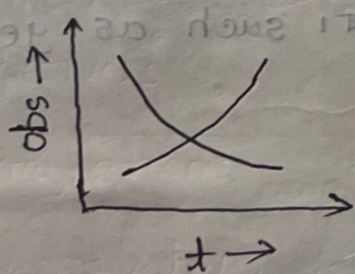


* mean বাড়ছে
(constant নয়)

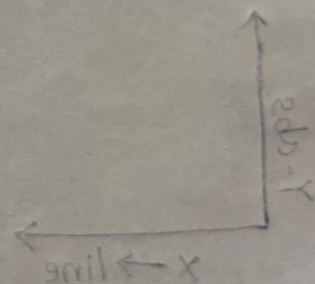
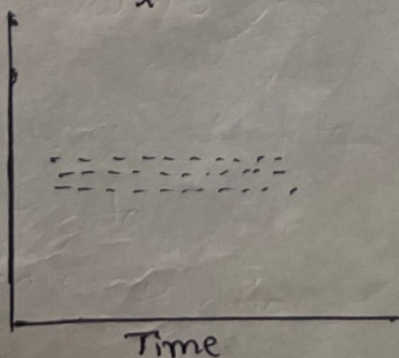
* trend → অন্বেষণ আপেক্ষিক graph upward or downward হতে পারে।

* upward trend এ mean, variance কমে increase করে।

* downward trend এ mean, var decrease করে।



independent of time
↑
stationary
obs



* No trend → mean and variance are approximately same over time, (constant)

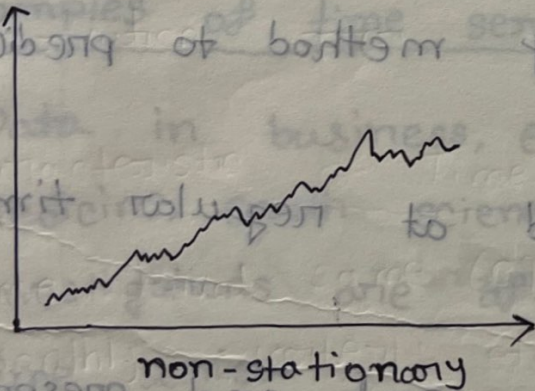
* যখন mean and variance constant হয় না তখন এগুলোকে constant রাখা যায় নিলে transformation করতে হবে।

* data set এর nature এর উপর depend করে transformation করে নিলে হবে।

* We have to transform trend line into no trend using different transformation methods to make mean, variance constant.

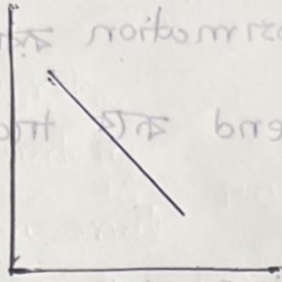
(ii) Seasonality/seasonal effect

* plot-এ কিছু অক্ষয় পর পর peak উপরে উঠে যাওয়াটা seasonality.



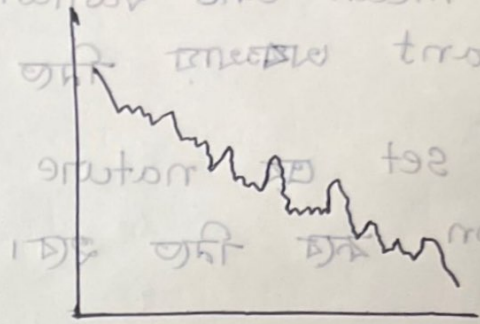
[seasonality remove করতে হবে]

* হঠাৎ করে উপরে উঠে যাওয়ায় বলা random fluctuation অথবা একে random variation বলে।



Non-stationarity

আছে



trend + non stationarity আছে

(remove করতে হবে)

trend remove করতে হবে।

Verbal definition

(i) both are non-stationary

(ii) both have trend line

(downward / decreasing)

(iii) only 2nd data has seasonal effect

→ A quantitative forecasting method to predict future values

→ Numerical data obtained at regular time intervals

→ Projections based on past and present observations

Two types of data:

(i) test data.

(ii) training data.

* Model develop করা হয় training data থেকে।

* Model check করা হয় test data থেকে।

Real life example

→ The demand of icecream, electricity increases at a particular season (summer)

Stationary / Non stationary

→ Fixed stage এ observe করা যায় যে stationary
তে নিয়ে যাওয়া হয়।

Examples of time series:

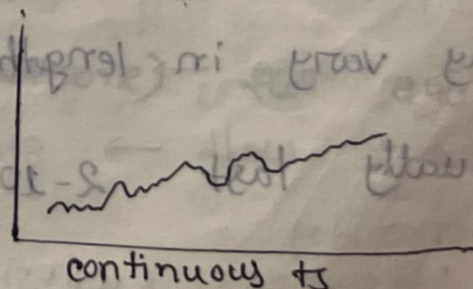
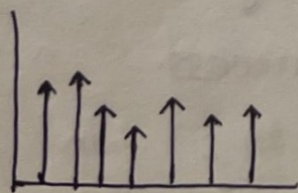
- Data in business, economics, engineering, environment, medicine, earth sciences, other areas of scientific investigations are often collected in the form of time series.

- Hourly temperature readings

- Daily stock prices.

- weekly traffic volume

- Annual growth rate.

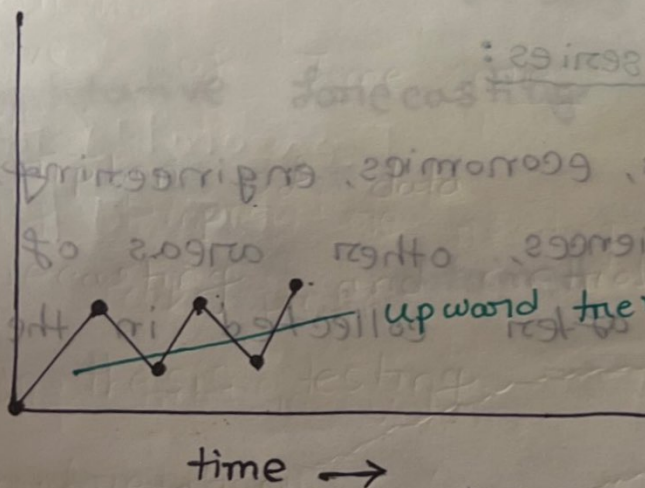


Time series component

- (1) Trend
- (2) Random
- (3) seasonal
- (4) cyclical

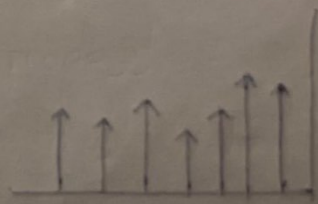
Trend Component:

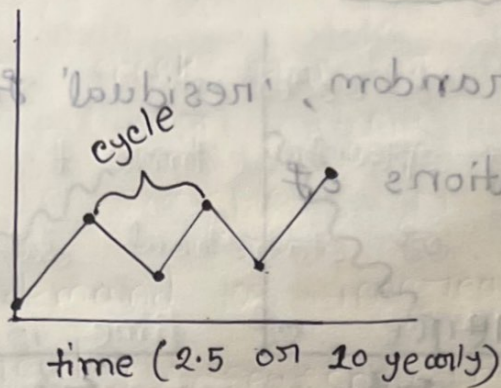
- Overall upward/downward movement
- Data taken over a period of years.
- time against measurement trend.



Cyclical component:

- Upward or downward swings
- May vary in length
- Usually last 2-10 years.

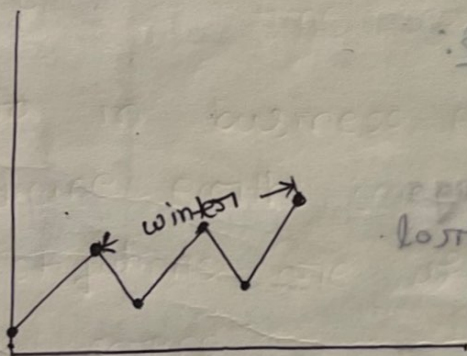




- Nature
- Accidents

Seasonal component

- Upward or downward swings.
- Regular pattern
- observed within one year.



time (monthly or quarterly) [within 1 year]

- ☐ Within one year a continuous peak & seasonal variation.
- ☐ Within 2, 5 or 10 (not one) years cyclical variation.

Random or Irregular component:

→ Erratic, nonsystematic, random, 'residual' functions.

→ Due to random variations of

- Nature
- Accidents

(time of day) error

→ Short duration and non-repeating (normally
ঘাড়ে but রাস উল্ল দেখে গেল)

* Time series is a special type of stochastic process.

Objective of Time series:

- ① Description of data
- ② Forecasting and control.
- ③ Hypothesis testing.
- ④ Interpretation.

Some important terms

→ Decomposition → ব্যাখ্যা করা

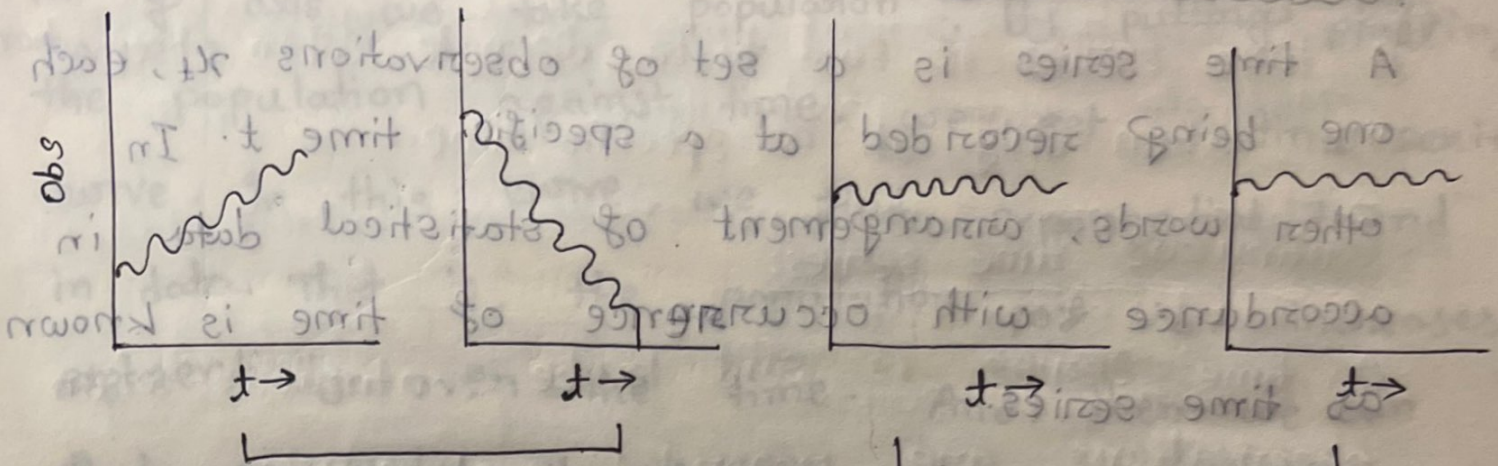
→ White noise → error forecasting শব্দ

→ Random walk → Stochastic process.

→ stationarity

• Skewness, kurtosis same.

- mean or variance doesn't change over time



stationary

Examples:

- Hourly temperature readings
- Daily stock prices
- Weekly traffic volume
- Annual growth rate

Practical Example:

suppose we have the data of population of the U.S.A. at 10 years intervals 1950-1960. that is a time series data. Now, for analyzing this data we

TS-1

Time series: analysis

A time series is a set of observations x_t , each one being recorded at a specific time t . In other words, arrangement of statistical data in accordance with occurrence of time is known as time series.

* Time series analysis is a quantitative forecasting method to predict future values, where numerical data are obtained at regular time intervals and projections are based on past and present observations.

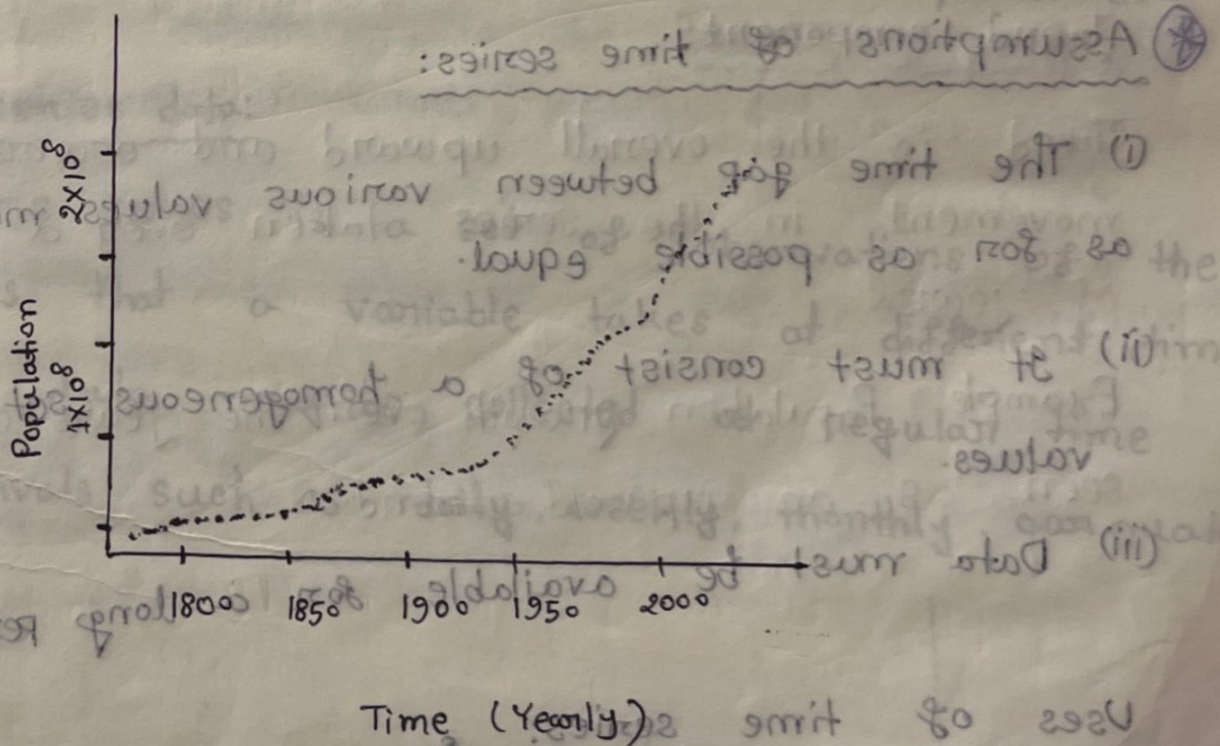
Examples:

- Hourly temperature readings.
- Daily stock prices.
- Weekly traffic volume
- Annual growth rate.

Practical Example:

Suppose, we have the data of population of the U.S.A at 10 years intervals 1790-1990. That is a time series data. Now, for analyzing this data we

have to plot it. In x axis we take "Year" and in y axis we take "population". By putting plotting the population against time, we get a time series curve. In this curve, we find exponential trend in data. That is, the population of U.S.A increases exponentially over the time. And then we can find or predict the future values of population. This is a practical example of time series analysis.



Time (Yearly)

Figure : 1

Types of time series:

There are two types of time series. such as

- (i) Discrete time series.
- (ii) Continuous time series.

Discrete time series:

A time series is said to be discrete, when observations are taken only at a specific time.

Continuous time series:

A time series is said to be continuous, when observations are recorded continuously over some time interval.

⊛ Assumptions of time series:

- (i) The time gap between various values must be as far as possible equal.
- (ii) It must consist of a homogeneous set of values.
- (iii) Data must be available for a long period.

Uses of time series:

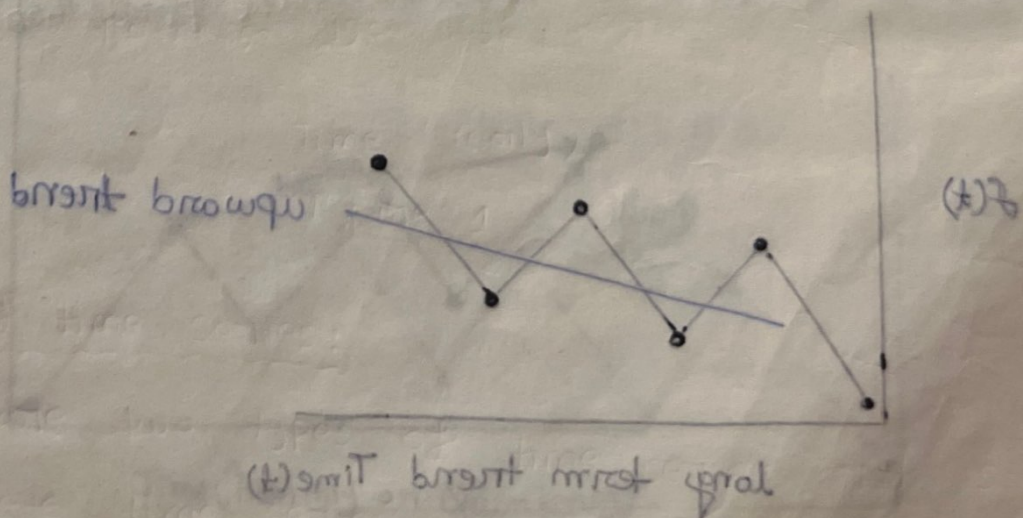
- (1) To determine the type and nature of the variations in data.
- (2) Helps in planning and forecasting.
- (3) Make plans for the future.

- (4) Helps to compare the actual current performance of accomplishments with the expected ones and analyze the causes of such variation, if any.
- (5) To predict or estimate or forecast the behavior of the phenomenon in future which is very essential for business planning.
- (6) To compare the changes in the values of different phenomenon at different times or places etc.

Time series data:

A time series is a set of observations of the values that a variable takes at different times.

Such data may be collected at regular time intervals such as daily, weekly, monthly, quarterly etc.



Time series Analysis

TS-2

What are the different time series components?
Illustrate them graphically.

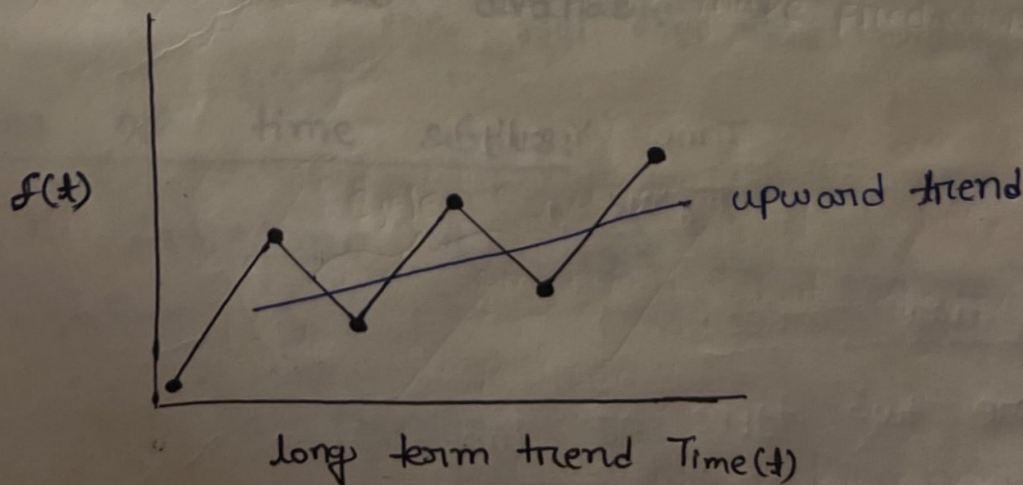
⇒ There are four time series components. They are:

- 1) Trend component
- 2) Seasonal component
- 3) Cyclical component
- 4) Random or irregular component.

Trend component:

Trend is the overall upward and downward movement in the series taken over a period of years.

Example: Population growth over the years can be seen as an upward trend.



- Population growth
- Agricultural production
- Export of manufactured goods.
- Prices of consumer goods.

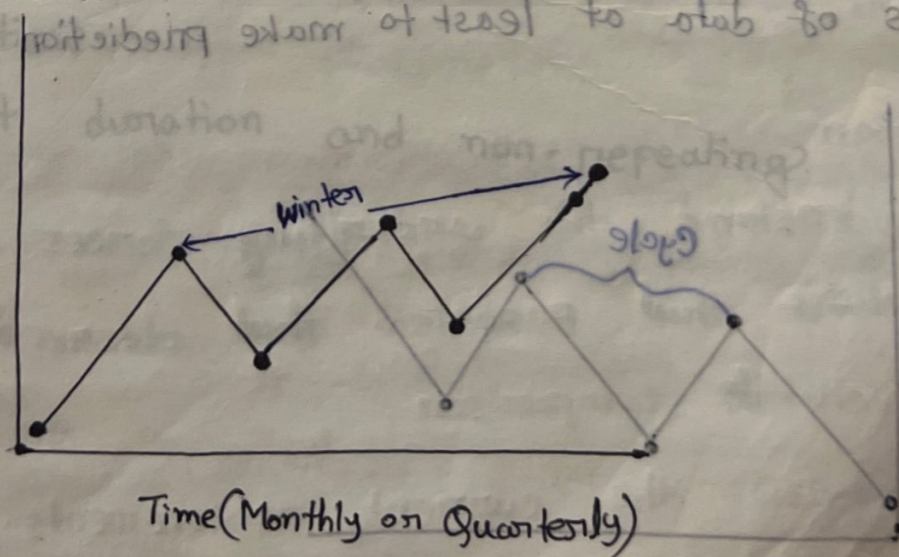
• A definite repeated movement of a regular periodic type, occurring over short duration is known as ~~season~~ seasonal variation.

Seasonal component:

Seasonal component is the regular patterns of upward or downward swings which is observed within 1 year.

- The time period may be a day, a week, a quarter or a year.

Ex: Sales of ice-cream increases during summer season.



Cyclical component:

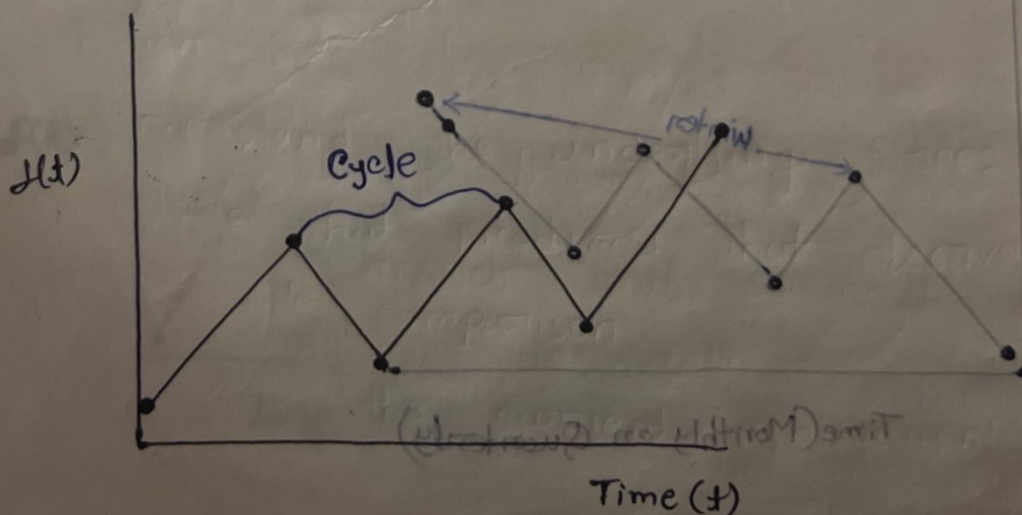
Cyclical component is the upward or downward swings which may vary in length and usually lasts 2-10 years.

Ex: 5 years of economic growth followed by 2 years of economic recession, 7 years of economic growth followed by 1 year of economic recession.

4 phases of cyclical component are:

- Prosperity
- Recession
- Depression
- Recovery

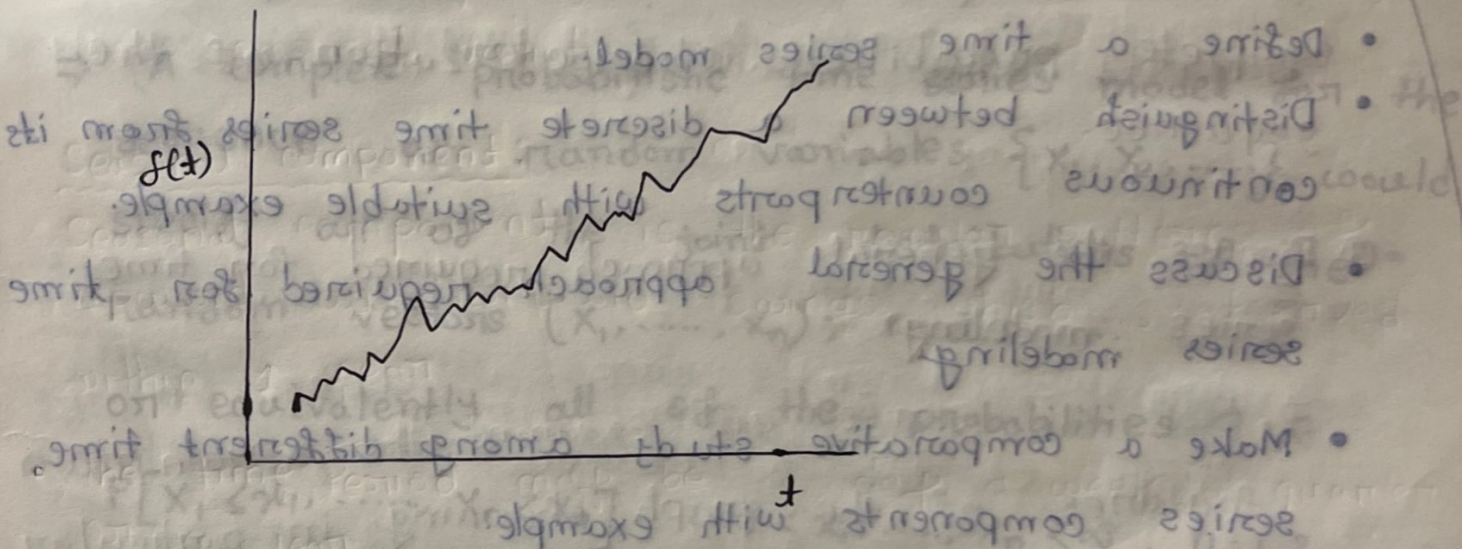
* Cyclical variations are difficult to forecast because of lack of historical data we need more than 30 years of data at least to make prediction about cycle.



Random or irregular component:

The variations which are happened due to uncertain and unusual cases (Natural disaster, accidents) is called random variation.

The time duration of this variation is short and it is non-repeating. It is purely random.



- The irregular fluctuations cannot be predicted.
- Shows a sudden rise or fall at certain instants of time.
- Short duration and non-repeating.

⊗ Mention different objectives of such analysis:

⇒ • Description of data.

• Forecasting and control.

• Hypothesis testing

• Interpretation

(TS-01 and TS-02)

• Define a time series model.

• Distinguish between a discrete time series from its continuous counterparts with suitable example.

• Discuss the general approach required for time series modeling.

• Make a comparative study among different time series components with example.

• What is meant by the second-order property of a time series model?

• Simulation

• To understand the underlying forces and structure that produced that observed data.

• To facilitate comparison.

• To review the current achievements.

① Define a time series model.
⇒ A time series model for the observed data $\{x_t\}$ is a specification of the joint distributions of a sequence of random variables $\{X_t\}$ of which $\{x_t\}$ is postulated to be a realization.

(2) What is meant by the second order property of a time series model?

⇒ A complete probabilistic time series model for the sequence of random variables $\{X_1, X_2, \dots\}$ would specify all of the joint distributions of the random vectors $(X_1, \dots, X_n)$; $n=1, 2, \dots$ or equivalently all of the probabilities

$$P[X_1 \leq x_1, \dots, X_n \leq x_n], -\infty < x_1, \dots, x_n < \infty, n=1, 2, \dots$$

such a specification is rarely used in time series analysis, since in general it will contain too many parameters to be estimated from the available data. Instead we specify only the first and second order moments of the joint distributions i.e. the expected values $E(X_t)$ and the expected products $E(X_{t+h}, X_t)$, $t=1, 2, \dots$, $h=0, 1, 2, \dots$ focusing on properties of the sequence $\{X_t\}$

that depend only on these. Such properties of $\{X_t\}$ are referred to as second order properties.

In the particular case where all the joint distributions are multivariate normal, the second-order properties of $\{X_t\}$ completely determine the joint distributions and hence give a complete probabilistic characterization of the sequence.

• ③ Discuss the general approach required for time series modeling.

⇒ • Plot the series and examine the main features of the graph, checking in particular whether there is

(a) a trend

(b) a seasonal component

(c) any apparent sharp changes in behaviour.

(d) any outlying observations

• Remove the trend and seasonal components to get stationary residuals. To achieve this goal it may sometimes be necessary to

apply a preliminary transformation to the data. There are several ways in which trend and seasonality can be removed.

(i) Estimating the components and subtracting them from the data.

(ii) Differencing the data i.e.,

replacing the original series $\{X_t\}$ by

$$(Y_t = X_t - X_{t-d}) \text{ for some positive integer } d.$$

- Choose a model to fit the residuals, making use of various sample statistics including the sample autocorrelation function.

- Forecasting will be achieved by forecasting the residuals and then inverting the transformation described above to arrive at forecasts of the original series $\{X_t\}$.

- An ~~exten~~ extremely useful alternative approach is to express the series in terms of its Fourier components.

* Our goal is to predict a series that typically is not deterministic but contains a random component. Stationarity helps us to do so.

TS-03

* What is meant by the stationarity of a time series? Do you think this is an important property while conducting a time series analysis?

Justify your answer.

⇒ A time series has stationarity if a shift in time doesn't cause a change in the shape of the distribution.

Basic properties of the distribution (mean, variance, covariance) are constant over the time.

Yes, I think stationarity is an important property while conducting a time series analysis. Because,

- It is easier to analyze.
- It makes the ability to understand, detect

and model the data with the procedures in time series analysis.

- The model assumptions are violated when nonstationary data is used.
- Most statistical forecasting methods are based on this assumption through the use of mathematical transformations.
- Stationarity is useful concept when thinking about estimation because it suggests that something is fixed across the sequence of random variables.

⊗ Suppose, you have a time series data which shows an increasing trend but no periodic component. How would you estimate this trend from this data? Explain

⇒ The model which contains no periodic component, the model will be of the form

$$X_t = m_t + Y_t$$

where, m_t is a slowly changing function known as the trend component and Y_t has zero mean

m_t is estimated by least square method.

In the least squares procedure we attempt to fit a parametric family of functions e.g.

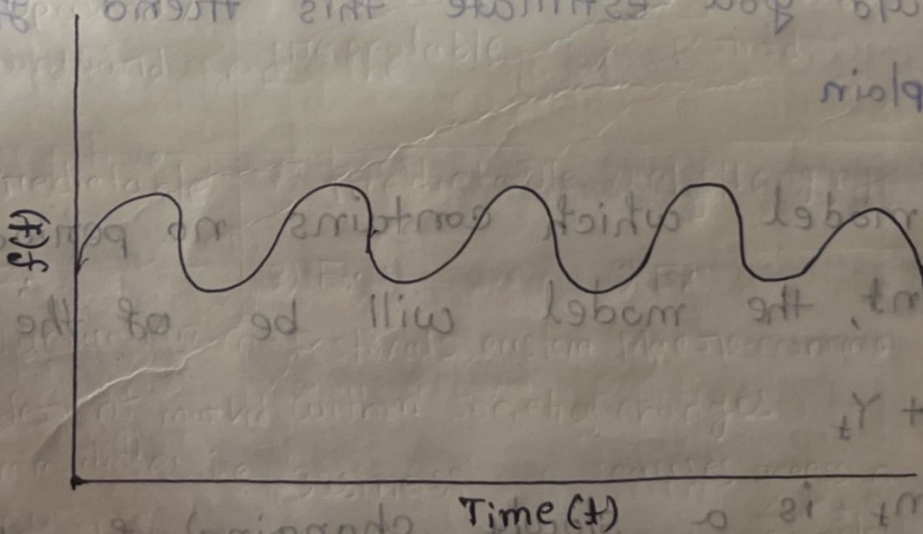
$$m_t = a_0 + a_1 t + a_2 t^2$$

to the data $\{x_1, x_2, \dots, x_n\}$ by choosing the parameters, (a_0, a_1, a_2) to minimize

$$\sum_{t=1}^n (x_t - m_t)^2$$

It can also be used to estimate higher order polynomial trends in the same way.

⊗ Sketch a time series plot which exhibits stationarity.

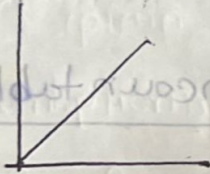
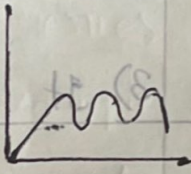
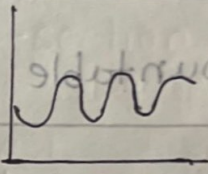
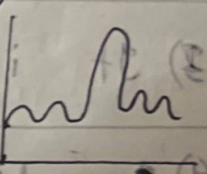


It is a time series plot which exhibits stationarity.

⊗ Distinguish between discrete time series and continuous time series with suitable example.

Discrete time series	Continuous time series
1) Observations are taken only at a specified time/ time interval.	1) Observations are recorded continuously over some time interval.
2) It is defined over a finite domain of sequence.	2) It is defined over a finite or infinite domain of sequence.
3) It is countable.	3) It is uncountable.
4) In which recorded observations are discrete.	4) In which recorded observations are continuous for a given interval.
5) Example: weekly stock market index	5) Example: the monthly accidental death data (1973-1978)

⊗ Make a comparative study among different time series components with example.

	Trend	Seasonality	Cyclical Fluctuations	Random or irregular
Duration	1) Long term	1) Short term	1) Long and short term	1) Short term
Denote	2) It is denoted by m_t	2) It is denoted by S_t	2) It is denoted by C_t	2) It is denoted by Y_t
Time	3) Fixed time interval	3) Fixed time interval	3) Not fixed time interval	3) Not fixed time interval
Visualization	4) 	4) 	4) 	4) 
Nature-1	5) Gradual	5) Swings between up or down	5) Repeating up and down	5) Erroneous or high fluctuation
Nature 2	6) Upward / downward	6) Pattern repeatable	6) No fixed period	6) Not repeatable
Prediction capability	7) Predictable	7) Predictable	7) Challenging	7) Challenging
	8) Ex: Fluctuation in different market commodities i.e gold, stock price, exchange rates etc	8) Ex: Festivals, custom within societies, Business Policies etc	8) Ex: Macro-economic changes causes recession and unemployment etc	8) Ex: Natural disaster, accidents etc

Autocorrelation Function (ACF)

- For data trend: ACF plot will exhibit slow decay as lag increases.
- For data with a substantial deterministic periodic component: Plot will exhibit similar behavior with the same periodicity.

Estimation and elimination of trend and seasonality:

$$X_t = m_t + S_t + Y_t$$

Classical decomposition model

where, $EY_t = 0$, $S_{t+d} = S_t$

and $\sum_{j=1}^d S_j = 0$

random noise component is stationary. If the seasonal and noise fluctuations appear to increase with the level of the process, then a preliminary transformation of the data is often used.

20-2T bmo 40-2T

After transformation: $\ln(x)$

The transformed data do not exhibit the increasing fluctuation with increasing level that was apparent in the original data. This suggests that the model is more appropriate for the transformed than for the original series.

Additive Decomposition:

$$Y_t = T_t + S_t + C_t + R_t$$

Assumption:

- 1) S_t , C_t , R_t operate with equal absolute effect irrespective of the trend.
- 2) C_t (and S_t) will have positive or negative values and the total of positive and negative values for any cycle will be zero (any year).
- 3) R_t will also have positive or negative values and in the long run $\sum R_t$ will be zero.

Multiplicative Decomposition:

$$Y_t = T_t \times S_t \times C_t \times R_t$$

$$\ln Y_t = \ln T_t + \ln S_t + \ln C_t + \ln R_t$$

Assumptions:

1) All the values are positive.

2) The geometric mean of S_t in a year, C_t in a cycle and R_t in a long term period are unity.

• To assess the degree of dependence in the data and to select a model for the data that

reflect a model for this, one of the important tools we use is the sample

autocorrelation function (sample ACF) of the data.

• Cross-sectional data is a collection of

observations for multiple subjects at a single

Tutorial-01 (TS-01 to TS-05)

(Note for exam)

Multiplicative Decomposition:

(TS-01) ① What do you mean by time series?

Answer: A time series is a set of observations

x_t , each one being recorded at a specific time t .

In other words, arrangement of statistical data in accordance with occurrence of time is known as time series.

(TS-01) ② Explain time series, cross-sectional and panel data with example.

Answer:

- Time series data is a collection of observations for a single subject at different time intervals.

For example: Max temperature, Humidity and wind (All three variables) in New York city (single entity) collected on first day of every year (multiple intervals of time)

- Cross-sectional data is a collection of observations for multiple subjects at a single point in time.

For example: Max temperature, Humidity and wind (all three variables) in New York city, SFO, Boston, Chicago (multiple entities) 1/1/2015 (single instance)

• Panel data is a type of data that professionals collect by observing particular variables over a period of time at a regular frequency.

For example: Panel data may comprise annual income information and the age of individuals over a nine year period.

(TS-01) ③ What is meant by a time series analysis?

Explain with suitable examples.

Answer:

Time series analysis is a quantitative forecasting method to predict future values, where numerical data are obtained at regular time intervals and projections are based on past and present observations.

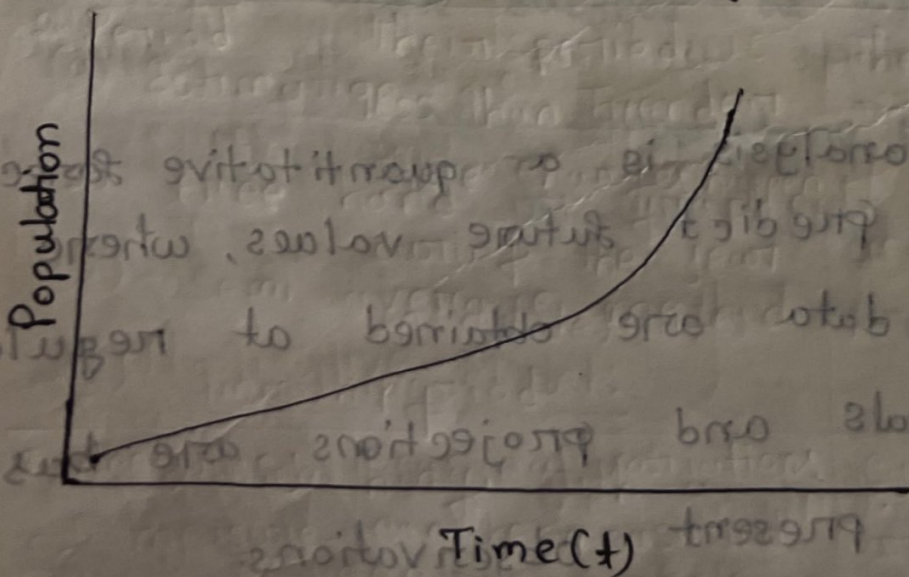
Practical Example:

Suppose we have a data of population of the USA at 10 years intervals 1790-1990, which is a time series data.

Now, for analyzing this data, we have to plot it. In x axis we take "Year" and in y axis we take "population". By plotting the population against time, we get a time series curve.

In this curve, we find exponential trend in the data. That means, the population of the USA increases exponentially over time.

Then we can find or predict the future values of the population using this information.



Other examples:

i) Hourly temperature reading.

ii) Daily stock prices.

iii) Weekly traffic volume.

iv) Annual growth rate.

(TS-02) ④ Mention different objectives of such analysis.

Answer:

1) Description of data.

2) Forecasting and control.

3) Hypothesis testing.

4) Interpretation.

5) Simulation.

6) To understand the underlying forces and structure that produced the observed data

7) To facilitate comparison.

8) To review the current achievements.

(TS-01) ⑤ Distinguish between a discrete time series from its continuous counterpart with suitable example.

Answer: Already noted.

(TS-02) ⑥ What are the basic differences between the seasonality and cyclical variation? Explain with examples.

Answer:

Basis of differences	Seasonal variations	Cyclical Fluctuations
1. Regularity	Both period and succession in seasonal variations are regular and their succession is definite.	Succession in cyclical fluctuations is fixed but change takes place in the period of each phase.
2. Period	Their period is less than one year. These changes happen during the year in a definite period.	The term in these cases is more than a year and is usually from 3 years to 10 or 12 years.
3. Accuracy of measurement	Seasonal variations can be measured more accurately.	Cyclical fluctuations are can not be measured very accurately.

Basis of Difference	Seasonal variations	Cyclical fluctuations
4. Causes of change	Seasonal variations are caused by changes season, traditions, fashions etc	Cyclical fluctuations are caused by currency expansion or contraction, decrease or increase in sale, sudden change in demand etc
5. Succession or nature of appearance	Seasonal variation is of different nature. They appear in every trade in different succession.	Cyclical fluctuations are of the same nature and appear in every trade uniformly
6. Method of measurement	They are measured in the following method: seasonal variation index method, variable mean method, range value ratio method and trend ratio method.	Cyclical method is used to measure cyclical fluctuations.
7. Example	Sales of ice-cream increases during summer season.	Macro-economic changes cause recession, unemployment etc.

(TS-02) ⑦ What is meant by deseasonalizing a time series?

Answer:

Seasonal adjustment or deseasonalization is a statistical method for removing the seasonal component of time series. It is usually done when wanting to analyse the trend, and cyclical deviations from trend of a time series independently of the seasonal components.

⑧ What are the different techniques to estimate and eliminate the trend component in the absence of seasonality.

Answer:

(a) Estimation of Trend:

1) Free hand or graphical method.

2) Semi-average method.

3) Smoothing

i) Smoothing with a finite moving average filter.

ii) Exponential smoothing.

(iii) Smoothing by elimination of high frequency components:

(iv) Polynomial fitting by principle of least square.

(b) Trend elimination by differencing:

This method has the advantage that it usually requires the estimation of fewer parameters and does not rest on the assumption of a trend that remains fixed throughout the observation period.

⑨ Explain the additive and multiplicative models of a time series stating clearly the assumptions and also discuss their relative merits.

Answer:

Additive model:

According to the additive model, a time series can be expressed as:

$$Y_t = T_t + S_t + C_t + U_t$$

Where Y_t is the time series value at time t ,
 T_t is the trend value at time t , S_t is the
seasonal variation at time t , C_t is the cyclic
variation at time t and U_t is the random
variation at time t .

Assumptions:

i) S_t , C_t , U_t operate with equal absolute effect
irrespective of the trend.

ii) C_t (and S_t) will have positive or negative
values and the total of the positive and
negative values for any cycle (and any

year) will be zero.
iii) U_t will also have positive or negative
values and in the long run $\sum U_t$ will be
zero.

Multiplicative Model:

According to the multiplicative model, a
time series can be expressed as

$$Y_t = T_t \cdot S_t \cdot C_t \cdot U_t$$

The multiplicative decomposition of a time series is same as the additive decomposition of the logarithmic values of the original time series. i.e.

$$\ln Y_t = \ln T_t + \ln S_t + \ln C_t + \ln U_t$$

Assumptions:

- i) All the values are positive.
- ii) The geometric mean of S_t in a year, C_t in a cycle and U_t in a long term period are unity.

Merits:

The additive model is useful when the seasonal variation is relatively constant over time. ①

The multiplicative model is useful when the seasonal variation increases over time. ②

⑩ If the estimated residuals have no trend and show smoothness, how would you quantify such feat. feature.

Answer:

Smoothness of the graph of a time series is generally indicative of the existence of some form of dependence among the observations.

Such dependence can be used to advantage in forecasting future values of the series.

We quantify such dependence by the autocorrelation function as a measure of dependence.

⑪ What is meant by the smoothness of a time series plot? How would you handle such smoothness in case of your data? Explain the concept with the visual presentation by a graph

Answer:

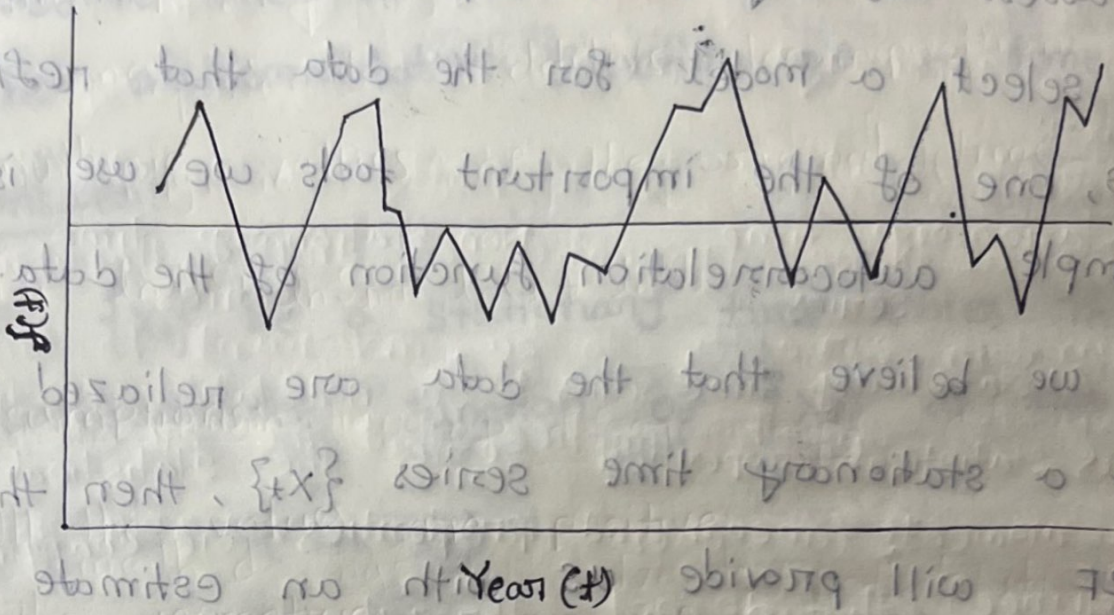


Figure: Residuals from fitting a line to the data

There are two interesting features of the graph of the residuals. The first is the absence of any discernible trend. The second is the smoothness of the graph.

Smoothness of the graph of a time series is generally indicative of the existence of some of dependence among the observations. This will be to our advantage, since dependence means in particular that past observations of the noise sequence can assist in predicting future values.

To assess the degree of dependence in the data and to select a model for the data that reflects this, one of the important tools we use is the sample autocorrelation function of the data.

If we believe that the data are realized values of a stationary time series $\{x_t\}$, then the sample ACF will provide us with an estimate of the ACF of $\{x_t\}$. This estimate may suggest which of the many possible stationary time series models is a suitable candidate for representing the dependence in the data.

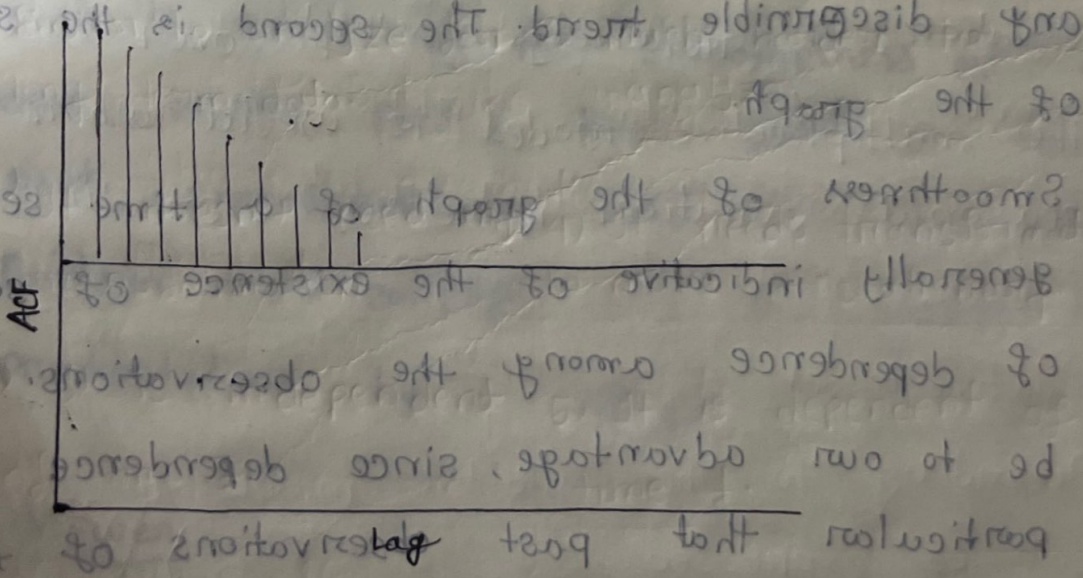


Figure: The sample autocorrelation function for the 'XYZ' residuals.

- (12) Define the autocovariance function (ACVF) and the autocorrelation function of a time series.

Answer:

Let $\{X_t\}$ be a stationary time series. The autocovariance function of $\{X_t\}$ at lag h is

$$\gamma_X(h) = \text{cov}(X_{t+h}, X_t)$$

The autocorrelation function (ACF) of $\{X_t\}$ at lag h is

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = \text{Corr}(X_{t+h}, X_t)$$

- (13) State and prove the basic properties of the ACVF of a time series.

Answer:

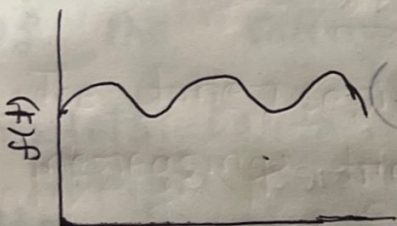
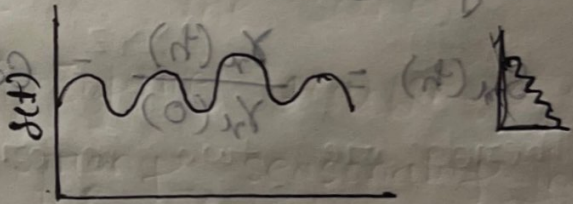
Linearity property of covariances:

If $EX^2 < \infty$, $EY^2 < \infty$, $EZ^2 < \infty$ and a , b , and c are any real constants, then

$$\text{cov}(aX + bY + c, Z) = a \text{cov}(X, Z) + b \text{cov}(Y, Z)$$

14) Distinguish between stationary time series and non-stationary time series.

Answer:

Stationary time series	Non-stationary time series
1) A stationary time series has statistical properties or moments (mean and variance) that do not vary in time	1) In nonstationary time series the statistical properties are changing through time
2) 	2) 
mean, variance are constant	mean, variance are not constant
3) Easier to analyze for long term data.	3) Much harder to analyze
4) It doesn't need any transformation.	4) It needs transformation to make it stationary.
5) It is independent of time t .	5) It is dependent of time t .

⑮ states the properties of a weakly and strictly stationary process.

Answer:

Weakly Stationary:

A time series $\{X_t, t=0, \pm 1, \pm 2, \dots\} = (X_t^T) < \infty$

is said to be weakly stationary if its ~~mean~~ mean function and covariance function is independent of time t .

Properties:

- i) $\mu_x(t)$ is independent of t and
- ii) $\gamma_x(t+h, t)$ is independent of t for each h .

It is usually termed as simply stationary.

Strictly stationary:

$\{X_t\}$ is a strictly stationary time series if

$$(X_1, \dots, X_n)' \stackrel{d}{=} (X_{1+h}, \dots, X_{n+h})'$$

for all integers h and $n \geq 1$.

Properties:

- i) The random variables X_t are identically distributed.
- ii) $(X_t, X_{t+h})' \stackrel{d}{=} (X_1, X_{1+h})'$ for all integers t and h .
- iii) $\{X_t\}$ is weakly stationary if $E(X_t^2) < \infty$ for all t .

iv) Weak stationary does not imply strict stationarity.

v) An IID sequence is strictly stationary. The strict stationarity property restricted the probabilistic properties of any collection of random variables to be invariant to time shifts, which implies that the probabilistic behavior of each random variable is the same across time.

⑩ Final Question - 2019

① The time series plot of Figure 1(a) shows an increasing trend with strong seasonality. There is no evidence of any cyclic behavior here. On the other hand, in Figure 1(b), there is no seasonality. But the graph suggests the probability of fitting a quadratic or exponential trend to the data.

The two time series plots are non-stationary.

Suggesting transformation: 1(a)

In Figure 1(a) there exist trend and seasonality. If the seasonal and noise fluctuations appear

to increase with the level of the process, then a preliminary transformation of the data is often used. Here, we see this characteristics in the plot. So, we can use logarithmic transformation and can see that the transformed data do not exhibit the increasing fluctuation with increasing level that was apparent in the original data.

② Zero mean model is inappropriate for the above two cases in figure 1(a) and (b)

Because the zero-mean models have a zero-mean and constant variance and shows no predictable trend or seasonality.

Mathematical model for 1(a):

$$X_t = m_t + S_t + Y_t$$

where, m_t is a slowly changing function known as a trend component.

S_t is a function with known period referred to as a seasonal component, and Y_t is a random noise component, that is stationary.

Mathematical model for 1(b)

$$X_t = m_t + Y_t ; t = 1, \dots, n$$

③ There are two interesting features of the graph of the residuals. The first is the absence of any discernible trend. The second is the smoothness of the graph.

Smoothness of the graph of a time series is generally indicative of the existence of some form of dependence among the observations.

Such dependence can be used to advantage in forecasting future values of the series.

Such pattern of the residuals would be very unlikely to occur if the residuals were observations of iid noise with zero mean.

(TS-01) ⑪ Write down the assumptions of time series.

Answer:

- 1) The time gap between various values must be as far as possible equal.
- 2) It must consist of a homogeneous set of values.
- 3) Data must be available for long period.

(TS-01) ⑧ Write down the uses of time series.

Answer:

- 1) To determine the type and nature of the variations in data.
- 2) Helps in planning and forecasting.
- 3) Make plans for the future.
- 4) Helps to compare the actual current performance of accomplishments with the expected ones and analyze the causes of such variation, if any.
- 5) To predict or estimate or forecast the behavior of the phenomenon in future which is very essential for business planning.

Trend elimination by differencing:

Instead of attempting to remove the noise by smoothing, we now attempt to eliminate the trend term by differencing. We define the lag-1 difference operator by ∇

$$\nabla X_t = X_t - X_{t-1} = (1-B)X_t$$

where B is the backward shift operator,

$$BX_t = X_{t-1}$$

Again,

$$\nabla^2 X_t = \nabla(\nabla(X_t))$$

$$= (1-B)(1-B)X_t$$

$$= (1-2B+B^2)X_t$$

$$= X_t - 2X_{t-1} + X_{t-2}$$

If the operator ∇ is applied to a linear trend function

then we obtain the constant function

$$\nabla m_t = m_t - m_{t-1}$$

$$= (a_0 + a_1 t) - (a_0 + a_1(t-1))$$

$$= a_0 + a_1 t - a_0 - a_1 t + a_1$$

$$= a_1$$

In the same way, any polynomial trend of degree k can be reduced to a constant by application of the operator ∇^k

For example, if $X_t = m_t + Y_t$; where $m_t = \sum_{i=0}^k a_i t^i$ and Y_t is stationary with mean zero, application of ∇^k gives

$$\nabla^k X_t = k! a_k + \nabla^k Y_t$$

where, $\nabla^k m_t = k! a_k$

a stationary process with mean $k! a_k$

⑩ Estimation and elimination of Both trend and Seasonality.

Answer:

① "Classical Decomposition", in which trend and/or seasonality are estimated and subtracted from the data to generate a noise sequence.

Explanation:

Classical decomposition model

$$X_t = m_t + S_t + Y_t ; t = 1, \dots, m$$

where, $EY_t = 0$, $S_{t+d} = S_t$

$$\text{and } \sum_{j=1}^d S_j = 0$$

- Suppose, we have observations $\{x_1, \dots, x_n\}$. The trend is first estimated by applying a moving average filter specially chosen to eliminate the seasonal component and to dampen the noise. If the period d is even, say $d=2q$ then we use,

$$\hat{m}_t = \frac{(0.5x_{t-q} + x_{t-q+1} + \dots + x_{t+q-1} + 0.5x_{t+q})}{d}$$

where, $q < t \leq n-q$

if the period d is odd, say $d = 2q+1$, then we use (simple moving average)

$$\hat{m}_t = (2q+1)^{-1} \sum_{j=-q}^q X_{t-j} ; \quad q+1 \leq t \leq n-q$$

• The second step is to estimate the seasonal component.

For each $k=1, \dots, d$, we compute the average w_k of the deviations

$$\{ (X_{k+jd} - \hat{m}_{k+jd}), \quad q < k+jd \leq n-q \}$$

since these average deviations do not necessarily sum to zero, we estimate the second seasonal component $\hat{s}_k$

$$\hat{s}_k = w_k - d^{-1} \sum_{i=1}^d w_i ; \quad k=1, \dots, d$$

and $\hat{s}_{k+d} = \hat{s}_k, \quad k > d$

• The deseasonalized data is then defined to be the original series with the estimated seasonal component removed. i.e.,

$$\hat{d}_t = X_t - \hat{s}_t, \quad t=1, \dots, n$$

- Finally, we reestimate the trend from the deseasonalized data $\{d_t\}$ using one of the methods such as polynomial fitting.
- In terms of this reestimated trend and the estimated seasonal component, the estimated noise series is then given by,

$$\hat{Y}_t = x_t - \hat{m}_t - \hat{S}_t \quad ; \quad t = 1, \dots, n$$

The reestimation of the trend can be done in (order to) have a parametric form for the trend that can be ~~extrapo~~ extrapolated for the purposes of prediction and simulation.

- ② "Differencing" in which trend and/or seasonal components are removed from the data by repeated differencing at one or more lags in order to generate a noise sequence.

Explanation:

The technique of differencing that we applied earlier to non-seasonal data can be adapted to deal with seasonality of period d by introducing the lag- d differencing operator ∇_d

defined by,

$$\nabla_d X_t = X_t - X_{t-d} = (1 - B^d) X_t$$

Applying the operator ∇_d to the model

$$X_t = m_t + S_t + Y_t$$

where $\{S_t\}$ has period d , we obtain

$$\nabla_d X_t = m_t - m_{t-d} + Y_t - Y_{t-d}$$

which gives a decomposition of the difference

$\nabla_d X_t$ into a trend component $(m_t - m_{t-d})$ and

a noise term $(Y_t - Y_{t-d})$. The trend $m_t - m_{t-d}$

can then be eliminated using the methods

already described, in applying a power of

the operator ∇_d

(20) Distinguish between seasonal variation and cyclical variation.

Answer:

Seasonal variation	Cyclical variation
① Involves short term Period	① Involves long duration
② There is no sequence	② A sequence is followed

Seasonal variation	Cyclical variation
③ Variations got some degree of regularity.	③ No regular variations are recorded.
④ Almost identical on same patterns arises every year.	④ Patterns completely depends upon market conditions.
⑤ Helps to predict demand of the product as per upcoming time	⑤ Helps to formulate policies as per market i.e, if there is opportunity, how to not take its benefit, but if there is threat how to avoid our weakness to overcome.
⑥ Lasts almost within a year	⑥ Usually lasts 2-10 years
⑦ It causes due to weather change, social customs, festivals etc	⑦ It causes due to 4 phases of business cycle (i) Prosperity (ii) Recession, (iii) Depression, (iv) Recovery
⑧ Examples: Sales of ice-cream increases during summer season.	⑧ Example: Macroeconomic changes cause recession and da unemployment

②1 The sample autocorrelation function:

To assess the degree of dependence in the data and to select a model for the data that reflects this one of the important tools we use is the sample autocorrelation function (sample ACF) of the data.

If we behave that the data are realized values of a stationary time series $\{X_t\}$, then the sample ACF will provides with an estimate of the ACF of $\{X_t\}$. This estimate may suggest which of the many possible stationary time series models are a suitable candidate for representing the dependence in the data.

Let $x_1, x_2, \dots, x_n$ be observations of a time series. The sample mean of $x_1, \dots, x_n$ is

$$\bar{x} = \frac{1}{n} \sum_{t=1}^n x_t$$

The sample autocovariance function is,

$$\hat{\gamma}(h) = \frac{1}{n-|h|} \sum_{t=1}^{n-|h|} (x_{t+|h|} - \bar{x})(x_t - \bar{x}), \quad -n < h < n$$

The sample autocorrelation function is,

$$\hat{\delta}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}, \quad -n < h < n$$

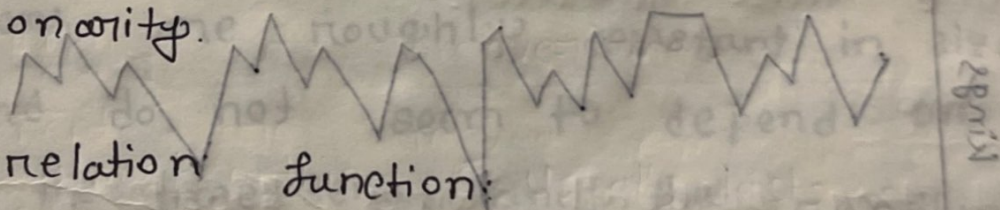
Remarks:

The sample autocovariance and autocorrelation functions can be computed for any data set $\{x_1, \dots, x_n\}$ and are not restricted to observations from a stationary time series.

- For data containing trend, $|\hat{\delta}(h)|$ will exhibit slow decay as h increases.
- For data with a substantial deterministic periodic component, $|\hat{\delta}(h)|$ will exhibit

Similar behaviour with the same periodicity.

Thus $\hat{\delta}(h)$ can be useful as an indicator of non-stationarity.



* Autocorrelation function:

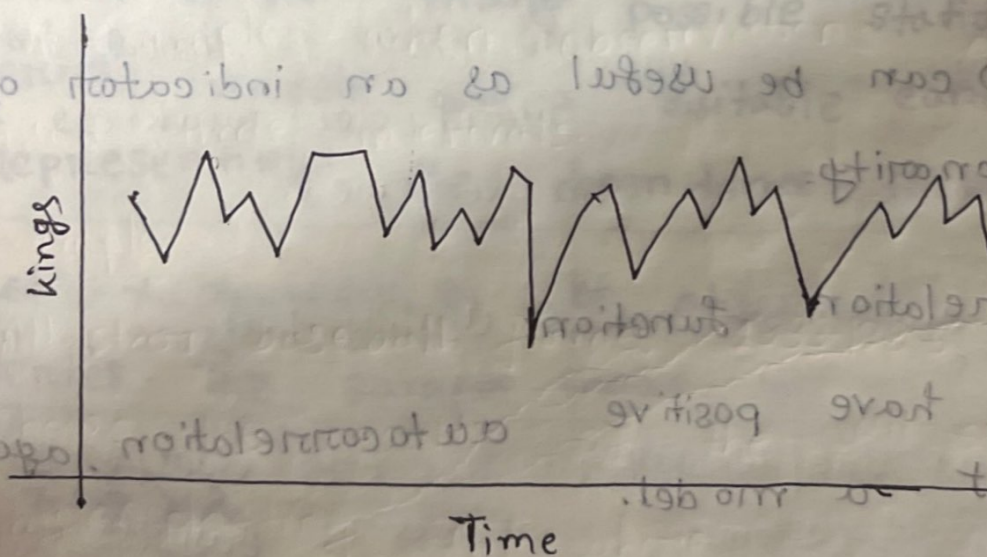
If we have positive autocorrelation, again we want a model.

If don't and it follows stationary process then our work is done.

Since the data are roughly constant in size over time.

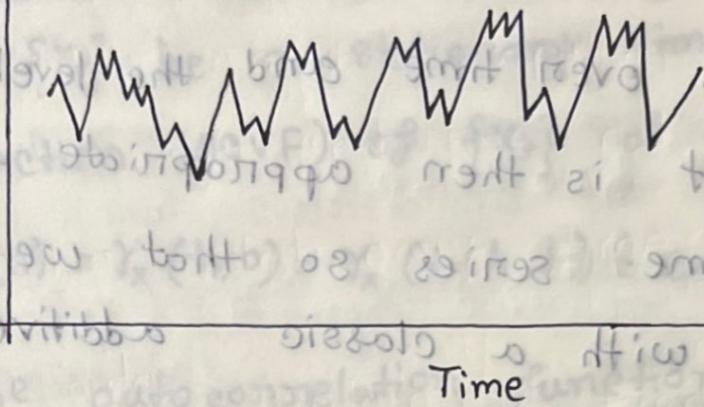
- The additive decomposition is the most appropriate if the magnitude of the seasonal fluctuations, or the variation around the trend-cycle, does not vary with the level of the time series. When the variation in the seasonal patterns or the variation around the trend-cycle, appears to be proportional to the level of the process, then a multiplicative decomposition is more appropriate.

①



At this points we could guess that this time series could be described using an additive model, since the random fluctuations in the data are roughly constant in size over time.

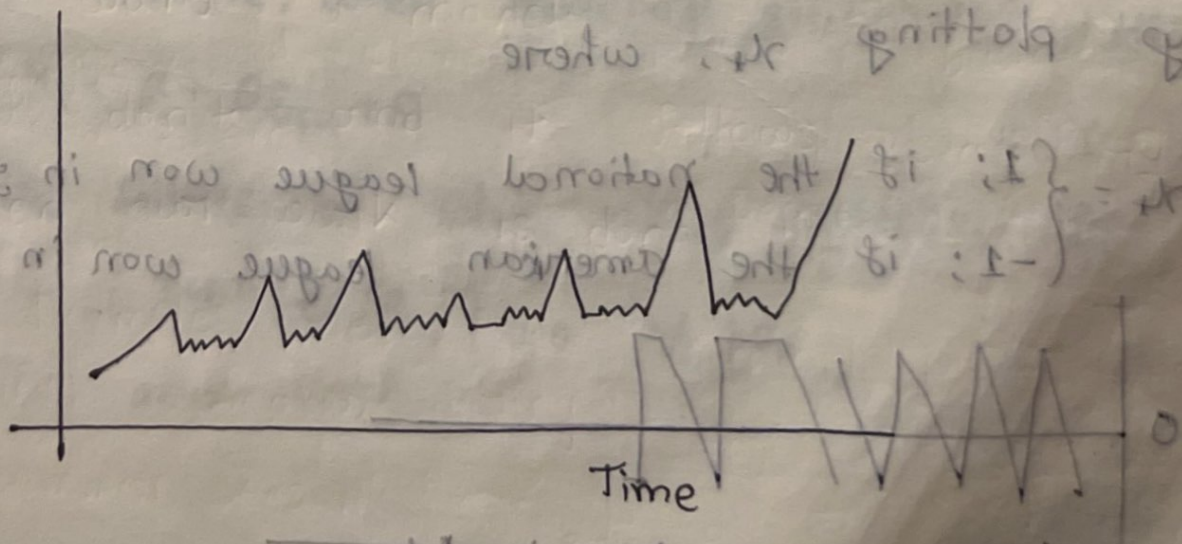
②



We can see from this time series that there is certainly some seasonal variation in the number of births per year; there is a peak in every summer and a trough in every winter.

Again it seems like this could be described using an additive model, as the seasonal fluctuations are roughly constant in size over time and do not seem to depend on the level of the time series, and the random fluctuations seem constant over time.

③



In this case, an additive model is not appropriate since the size of the seasonal and random fluctuations change over time and the level of the time series. It is then appropriate to transform the time series so that we can model the data with a classic additive model.

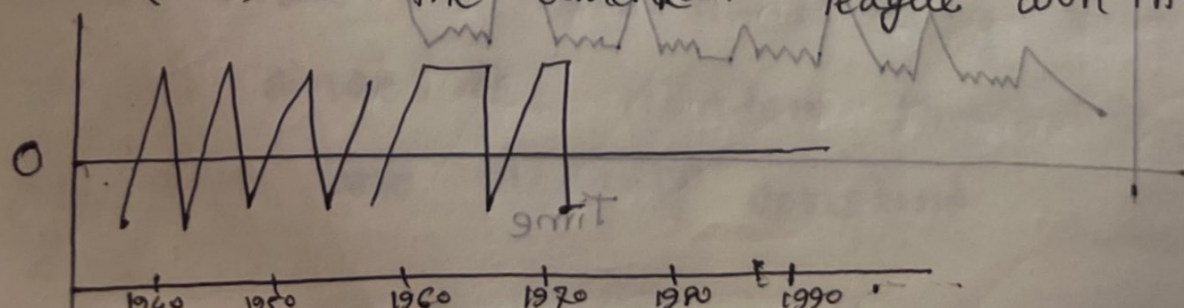
Discrete time series:

A discrete time series is one in which the set of times at which observations are made is a discrete set, as is the case, for example, when observations are made at fixed time intervals.

Example:

All star baseball games, 1933-1995. Figure 12 shows the results of the all star games by plotting x_t , where

$$x_t = \begin{cases} 1; & \text{if the national league won in year } t \\ -1; & \text{if the american league won in year } t \end{cases}$$



Stationary time series: $\{X_t, t = 0, \pm 1, \dots\}$

Let $\{X_t\}$ be a stationary time series. The autocovariance function (ACVF) of $\{X_t\}$ at lag h is,

$$\gamma_x(h) = \gamma_x(h, 0) = \gamma_x(t+h, t) = \text{cov}(X_{t+h}, X_t)$$

The autocorrelation function (ACF) of $\{X_t\}$ at lag h is,

$$S_x(h) = \frac{\gamma_x(h)}{\gamma_x(0)} = \text{corr}(X_{t+h}, X_t)$$

Here, $\gamma_x(\cdot)$: autocovariance function

$\gamma_x(h)$: value of $\gamma_x(\cdot)$ at lag h

White Noise:

(Samples could follow different probability distribution)

It is usually used to describe the behaviour of an error term in a model.

For example, when modeling stationary time series:

$$X_t = \alpha + \theta \varepsilon_{t-1} + \varepsilon_t$$

ε_t follows a white noise process, if

$$\text{corr}[\varepsilon_t, \varepsilon_{t-j}] = 0 \quad \forall j \neq 0$$

IID noise: (special case of white noise)

- Each sample has the same probability distribution.

$$\{X_t\} \sim \text{IID}(0, \sigma^2)$$

$$\gamma_x(t+h, t) = \begin{cases} \sigma^2 & \text{if } h=0 \\ 0 & \text{if } h \neq 0 \end{cases}$$

Definition:

Let $\{X_t\}$ be a time series with $E(X_t^2) < \infty$

The mean function of $\{X_t\}$ is,

$$\mu_x(t) = E(X_t)$$

The covariance function of $\{X_t\}$ is,

$$\begin{aligned} \gamma_x(\pi, s) &= \text{cov}(X_\pi, X_s) \\ &= E[(X_\pi - \mu_x(\pi))(X_s - \mu_x(s))] \end{aligned}$$

for all integers π and s

Basic properties of $\gamma(\cdot)$:

$$\gamma(0) \geq 0,$$

$$|\gamma(h)| \leq \gamma(0) \text{ for all } h,$$

and $\gamma(\cdot)$ is even, i.e.,

$$\gamma(h) = \gamma(-h) \text{ for all } h.$$

Proof:

The first property is simply the statement that $\text{var}(X_t) \geq 0$

The second is an immediate consequence of the fact that correlations are less than or equal to 1 in absolute value (or the Cauchy-Schwarz inequality).

and the third is established by observing that

$$\gamma(h) = \text{cov}(X_{t+h}, X_t)$$

$$= \text{cov}(X_t, X_{t+h})$$

- Autocovariance functions have another fundamental property, namely that of non-negative definiteness.

A real valued function k defined on the integers is non-negative definite if

$$\sum_{i,j=1}^n a_i k(i-j) a_j \geq 0$$

for all positive integers n and vectors a as $(a_1, \dots, a_n)$ with real valued components a_i

Smoothing with a finite moving average filter:

The moving average provides us with the estimates

$$\hat{m}_t = (2q+1)^{-1} \sum_{j=-q}^q X_{t+j}; \quad q+1 \leq t \leq n-q$$

Since, X_t is not observed for $t \leq 0$ or $t > n$

we can't use it for $t \leq q$ or $t > n-q$

White noise:

If we doing any time (series modeling) project, and we want to make predictions in the future, the moment we realize that a particular time series is white noise, it means that it is impossible to forecast values for that series in the future.

- Constant mean
- Constant variance
- There is no autocorrelation.

(That means the value of the time series at a particular periode has no relation with the values in the past time period.)

It is impossible to make predictions for a white noise because, there is no patterns in our time series data. (It's completely random).

Here, samples would follow different probability distribution.

IID noise:

The simplest model for a time series X is one in which there is no trend or seasonality and in which the observations are simply IID random variables with zero mean.

Every IID $(0, \sigma^2)$ sequence is WN $(0, \sigma^2)$ but not conversely.

$$\{X_t\} \sim \text{IID}(0, \sigma^2)$$

$$Y_x(t+h) - Y_x(t) = 0$$

$$Y_x(t+h, t) = \begin{cases} \sigma^2 & \text{if } h=0 \\ 0 & \text{if } h \neq 0 \end{cases}$$

[each sample has the same probability distribution]

TS-07

Preliminary transformation is not original transformation. It use to remove the jumping stages. (more stable form পাৰে)

Time series হয় $\rightarrow$ 2nd order moment finite হ'ল,

Strong / strict stationary time series

$$\{X_t, t = 0, \pm 1, \dots\}$$

X_1

X_2

X_3

$\vdots$

X_n

X_{t+h}

X

X_{n+h}

Probabilistic behaviour কে restrict কৰে দেয় কয়ন এখান
IID হ'ব হ'ব।

এটা নিয়ে কাজ হ'ব কৰে
আজ এখান যদি IID হ'ব

যায়, series IID হ'ব যায়, analyst হিচাবে আত্ম
Interest tell কৰো না। to step forward - কয়ন
এটা probabilistic behaviour টা loss কৰে।

Change of variation না থাকে = তখন analysis
কৰে more informative result পাওঁ যা না।

$X_t = X_t$ - Indexed by time

Complete time series model এ যতগুলো observation থাকে ঠিক ততগুলো parameters involve করতে হয়, ফলে একটি একটি tedious job. ফলে unnecessary complexity arise করে। আর দু'লম্বাছাড়া gain তেমন ফেরা নেই। তাই এই model কে অস্বাভাবিক avoid করা হয়।

IID noise (no dependency)

autocorrelation লাগবে।

But multicollinearity থাকবে।

এটিকে remove করতে হবে। (Treatment use করে)

Binary process তখন হয় যখন random variable টা Discrete হবে এবং তার outcome হবে দু'টো। Number of trial fixed থাকবে হবে।

Probability of success trial to trial same থাকবে। Discrete time series এ এটি থাকবে।

Random walk: (with drift and without drift):

Time এর বিভিন্ন change of combination এর কারণে এটি ঘটে থাকে।

$$S_t = X_1 + X_2 + \dots + X_t, \quad t = 1, 2,$$

$$S_0 = 0$$

$\{X_t \rightarrow \text{IID noise}$

A random walk is one in which future steps or directions cannot be predicted on the basis of past history.

TS-10 (Note)

Testing the estimated noise sequence:

After finding stationary residuals, the next step is to model the estimated noise sequence.

- If there is no dependence among these residuals, this means that the residuals are the observations of independent random variables.
- If there is significant dependence among the residuals, then we need to look for a more complex stationary time series model for the ~~model~~ noise that accounts such dependence.
- Dependence means that past observations of the noise sequence can assist in predicting future values.

H_0 : The estimated residuals are iid

(1) The sample autocorrelation function:

For large n , the sample autocorrelations of an iid sequence $Y_1, \dots, Y_n$ with finite variance are approximately iid with distribution $N(0, 1/n)$

Hence, if $y_1, \dots, y_n$ is a realization of such an iid sequence, about 95% of the sample autocorrelations should fall between the bounds $\pm 1.96/\sqrt{n}$

If we compute the sample autocorrelations up to lag 40 and find that more than two or three values fall outside the bounds, or one value falls far outside the bounds, we therefore reject the null hypothesis.

(2) The portmanteau Test:

Instead of checking to see whether each sample autocorrelation $\hat{\rho}(j)$ falls inside the bounds, it is also possible to consider the single statistic,

$$Q = n \sum_{j=1}^h \hat{\rho}(j)^2 \sim \chi^2_{(h)}$$

n = number of random sample observations

(5) $\hat{\rho}(j)$ = sample autocorrelation

h = number of lags we use

If $Q > \chi^2_{(1-\alpha)}(h)$, reject H_0

Some modification are there, formulated by Ljung and Box (1978) in which Q is replaced by

$$QLB = n(n+2) \sum_{j=1}^h \hat{\rho}(j)^2 / (n-j)$$

whose distribution is better approximated by chi-square distribution with h degrees of freedom

(3) The turning Point Test:

If T is the number of turning points of an iid sequence of length n , then the expected value of T is,

$$\mu_T = E[T] = \frac{2(n-2)}{3}$$

$$\sigma_T^2 = \text{Var}[T] = \frac{16(n-29)}{90} \quad [\text{variance of } T]$$

For large n ,

T is approximately $N(\mu_T, \sigma_T^2)$

If $|T - \mu_T| / \sigma_T > \Phi_{1-\alpha/2}$, reject H_0

• A large value of $(T - \mu_T)$ indicates that the series is fluctuating more rapidly than

expected for an iid sequence

- A value of $(T - \mu_T)$ much smaller than zero indicates a positive correlation between neighbouring observations

(4) The difference sign test:

For this test,

S = Number of times the difference series

$X_i - X_{i-1}$ is positive. $[i = 1, 2, \dots, n]$

For an iid sequence,

$$\mu_S = E[S] = \frac{n-1}{2}$$

$$\sigma_S^2 = \text{var}[S] = \frac{n+1}{12}$$

For large n ,

S is approximately $N(\mu_S, \sigma_S^2)$

A large positive (negative) value of $(S - \mu_S)$ indicates the presence of increasing (decreasing) trend in the data.

We therefore reject the assumption of no trend in the data if

$$|S - \mu_S| / \sigma_S > \Phi_{1-\alpha/2}$$

(5) The rank test:

For this test,

P = Number of pairs (i, j) such that $y_j > y_i$ and

$$j > i, [i = 1, 2, \dots, n-1]$$

For an iid sequence,

$$\mu_P = \frac{1}{4} n(n-1)$$

$$\sigma_P^2 = n(n-1)(2n+5)/72$$

For large n ,

P is approximately $N(\mu_P, \sigma_P^2)$

A large positive (negative) value of $(P - \mu_P)$ indicates the presence of an increasing (decreasing) trend in the data.

Reject H_0 : if $|P - \mu_P| / \sigma_P > \Phi_{1-\alpha/2}$

(6) Fitting an autoregressive model:

Fit an autoregressive model to the data using the Yule-Walker algorithm and choose the order that minimizes the AIC statistic. A selected order equal to zero suggests that the data is a white noise.

(7) Checking for normality:

- Histogram of Residuals
- Q-Q plot

Autoregressive models:

Features:

- No consistent trend
- No seasonality
- No obvious outliers
- Difficult to judge whether the variance is constant or not

Then AR(1) model might be appropriate.

Justification:

Plot values of the series against lag 1 values of the series.

t	Y_t	Y_{t-1}
1	13	-
2	14	13
3	8	14
4	10	8
5	16	10

Regress Y_t on Y_{t-1} . If it shows a moderately strong relationship, then the AR(1) model might be appropriate.

* A time series $\{X_t\}$ is said to be AR(p) if autoregressive of order (p) on AR(p) if $\phi(z) \neq 0$ and a moving average process of order q on MA(q) if $\theta(z) \neq 0$.

Definition:

An autoregressive (AR) model is when a value from a time series is regressed on previous values from that same time series.

For example:

Ψ_t on Ψ_{t-1}

$$\Psi_t = \beta_0 + \beta_1 \Psi_{t-1} + \varepsilon_t$$

$$\text{or, } X_t = \phi X_{t-1} + z_t$$

AR(1) or First order autoregressive model. —

$$\Psi_t = \beta_0 + \beta_1 \Psi_{t-1} + \beta_2 \Psi_{t-2} + \varepsilon_t$$

or,

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + z_t$$

AR(2) or second order autoregressive model

AR(p) model:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + z_t$$

where $\{z_t\} \sim WN(0, \sigma^2)$

MA(q) model:

$$X_t = \theta_1 z_{t-1} + \theta_2 z_{t-2} + \dots + \theta_q z_{t-q} + z_t$$

where $\{z_t\} \sim WN(0, \sigma^2)$

ARMA Model:

- The merger between $AR(p)$ and $MA(q)$ models.
 - Modelling financial time series.
 - Explains the momentum and mean reversion effects often observed in trading markets.
- +
- Captures the shock effects observed in the white noise terms.
- ARMA model attempts to capture both of these aspects of an AR and a MA model.
- It predicts the future values based on both the previous values and errors.
 - Thus, ARMA has better performance than AR and MA models alone.

ARMA(1,1)

$$X_t - \phi_1 X_{t-1} = z_t + \theta_1 z_{t-1}$$

ARMA(p,q)

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = z_t + \theta_1 z_{t-1} + \dots + \theta_q z_{t-q}$$

$\Phi(\cdot)$ and $\Theta(\cdot)$ are the p th and q th - degree polynomials

$$\Phi(z) = 1 - \Phi_1 z - \dots - \Phi_p z^p$$

$$\Theta(z) = 1 + \Theta_1 z + \dots + \Theta_q z^q$$

ARMA Model is a good example of the linear process.

Exponential Smoothing:

For any fixed $\alpha \in [0, 1]$, the one sided moving averages $\hat{m}_t$, $t = 1, 2, \dots, n$, defined by the recursions.

$$\hat{m}_t = \alpha x_t + (1 - \alpha) \hat{m}_{t-1}, \quad t = 2, \dots, n \quad (1)$$

$$\text{and, } \hat{m}_1 = x_1 \quad (2)$$

Specify the value of α . Application of (1) and (2)

is often referred to as exponential smoothing, since the recursions imply that for $t \geq 2$.

$$\hat{m}_t = \sum_{j=0}^{t-1} \alpha (1 - \alpha)^j x_{t-j} + (1 - \alpha)^{t-1} x_1$$

a weighted moving average of $x_t, x_{t-1}, \dots$, with weights decreasing exponentially (except for the last one).
Moving average vs exponential smoothing.

- ① In moving averages, the past observations are weighted equally. On the other hand, exponential smoothing assigns exponentially decreasing weights as the observation get older.
- ② The moving average removes noise using the mean value of data points, but exponential smoothing removes more noise from the data according to smoothing factor.
- ③ Exponential smoothing is less sensitive to outliers and noise, as it smooths them out with the weighted average than moving average.
- ④ Exponential smoothing is biased towards the past data, especially when the smoothing factor is low. This means that it may not react quickly enough to the new information or changes in the data.
- ⑤ Exponential smoothing may also underestimate or overestimate the forecast depending on whether the data is trending up or down.

(6) If cyclical variations presents in time series, then we use moving average.

(7) Exponential smoothing can be done using even two data points but to start the moving average we are required to set the window size which needs to be 'more than three' for performing actual smoothing.

Random Walk

A tool called a random walk can help you understand the predictability of our time series forecast problem.

- A random walk is not a sequence of random numbers.
- It is different from a list of random numbers because the next value in the sequence is a modification of the previous value in the sequence.
- The process used to generate the series forces dependence from one time step to the next.

This dependence provides some consistency from

step to step rather than the large jumps that a series of independent random numbers provides.

- Examples: Price of a security on the stock market.

- We would expect a strong autocorrelation with previous observation and a linear fall off from there with previous lags values.

- A random walk is nonstationary.

- We can make the random walk stationary by taking difference but we observe that we have no information to work here. Other than a series of random moves, there is no structure to learn.

- No significant relationship between the lagged observations.

- All correlations are small, close to zero and below the 95% and 99% confidence interval (beyond a few statistical lags).

- A random walk is unpredictable.

The best prediction we could make to use the observation at previous time step as what will happen in the next time step.

Simple, the next time step will be a function of the prior time step.

(Naive forecast or a persistence model)

Time series is a random walk if:

- It shows a strong temporal dependence that decays linearly or in a similar pattern.
- It is nonstationary and making it stationary shows no obviously learnable structure in the data
- The persistence model provides the best source of reliable predictions.

$$Y_t = \delta + Y_{t-1} + \epsilon_t$$

if $\delta = 0$, then the random walk is said to be without drift

if $\delta \neq 0$, then the random walk is with drift

$$Y_t = Y_0 + at + \sum_{i=1}^n \epsilon_i$$

on, $Y_t = Y_0 + \delta t + \sum_{i=1}^t \epsilon_i$, $t > 0$ prediction
 then, observation at previous time step as well

$$E(Y_t) = Y_0 + \delta t$$

$$\text{Var}(Y_t) = \sigma^2 t$$

$$\text{Cov}(Y_i, Y_j) = 0 \text{ for } i \neq j$$

The variance values are not constants but vary with time t , so this type of time series is non-stationary. Also, the mean values are constant only for a random walk without drift.

$$E(Y_t) = Y_0$$

$$\text{Var}(Y_t) = \sigma^2 t$$

Deterministic trend

$$Y_t = \delta t + \epsilon_t \rightarrow E(Y_t) = \delta t$$

$$\text{Var}(Y_t) = \sigma^2$$

Stochastic trend

$$Y_t = \delta + Y_{t-1} + \epsilon_t$$

$$E(Y_t) = \delta + Y_{t-1}$$

$$\text{Var}(Y_t) = \sigma^2 t$$

With drift

$$* Y_t = \delta + Y_{t-1} + \varepsilon_t$$

$$= \delta + \delta + Y_{t-2} + \varepsilon_{t-1} + \varepsilon_t + \delta = Y_t$$

$$= \delta + \delta + \delta + Y_{t-3} + \varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t + \delta = Y_t$$

$$\therefore Y_t = \delta t + Y_0 + \sum_{i=0}^{t-1} \varepsilon_{t-i} = Y_t$$

$$E(Y_t) = \delta t$$

$$\text{var}(Y_t) = t\sigma^2$$

$$Y_t = Y_{t-1} + \varepsilon_t$$

$$= Y_{t-2} + \varepsilon_{t-1} + \varepsilon_t$$

$$\therefore Y_t = Y_0 + \sum_{i=0}^{t-1} \varepsilon_{t-i}$$

$$E[Y_0] = 0$$

$$\Rightarrow E[Y_t] = 0$$

$$\text{var}(Y_t) = t\sigma^2 = f(t)$$

Summary

Time HW

Random walk with drift $x_3 + x_{1-3} + x = x$ *

$$y_t = \alpha + y_{t-1} + \varepsilon_t \quad x_3 + x_{1-3} + x_{5-3} + x + x =$$

$$\Rightarrow y_t = \alpha + \alpha + y_{t-2} + \varepsilon_{t-1} + \varepsilon_t \quad x + x + x =$$

$$\Rightarrow y_t = \alpha + \alpha + \alpha + y_{t-3} + \varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t = x$$

$$\therefore y_t = \alpha + y_0 + \sum_{i=0}^{t-1} \varepsilon_{t-i}$$

$$x + x = (x) \in$$

$$x + x = (x) \text{ root}$$

$$\Rightarrow y_t = y_0 + \alpha t + \sum_{i=0}^{t-1} \varepsilon_{t-i}$$

$$x + x = x$$

Random walk with drift

$$x + x + x =$$

$$\text{mean} = E(y_t) = \alpha t$$

$$\text{variance} = \text{Var}(y_t) = t\sigma^2 \quad x + x + x = x$$

Random walk without drift:

$$0 = [x] \in$$

$$y_t = y_{t-1} + \varepsilon_t$$

$$0 = [x] \in$$

$$(x) = -x = (x) \text{ root}$$

$$\Rightarrow y_t = y_{t-2} + \varepsilon_{t-1} + \varepsilon_t$$

$$\Rightarrow y_t = y_0 + \sum_{i=0}^{t-1} \varepsilon_{t-i} \quad [y_0 + \sum_{i=1}^n \varepsilon_i]$$

$$\therefore E(y_t) = 0$$

$$\text{Var}(y_t) = t\sigma^2$$

$$\therefore E(y_0) = 0$$

Deterministic trend:

$$Y_t = \alpha_t + \epsilon_t$$

$$E(Y_t) = \alpha_t$$

$$\text{var}(Y_t) = \sigma^2$$

Stochastic trend:

$$Y_t = \alpha + Y_{t-1} + \epsilon_t$$

$$E(Y_t) = \alpha$$

$$\text{var}(Y_t) = t\sigma^2$$

Drift — unexpected and undocumented changes to data structure

• A trend is a function of time but a drift is essentially just an intercept.

• ACF calculates the similarity of the values of a time series with its own lagged values.

The series is lagged step by step and the correlation between the original and the shifted series is calculated. This is done up to a max lag value.

• ACVF calculated the covariance of a time series with itself.

ACF — mean sumes the linear relationship between a time series and its past obs at different lags.

ACVF — covariance (how the values of a time series are related to each other as the time lag between the observational changes.

Second order property is important because:

- Help to check for stationarity in the data.
- Play a vital role in selecting appropriate models. in time series, having an accurate model allows
- Having an accurate model allows for better forecasting performance.
- Used to identify heteroscedasticity which refers to the varying levels of variability in a time series.
- Essential for model model diagnostic checking.
- Helps to identify outliers or unusual observed observation in time series data.