

1st Tutorial

Time series: 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.

Discrete time series: A discrete time series is one in which the set T_0 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: Daily stock prices or monthly temperature reading.

Continuous time series: Continuous time series are obtained when observations are recorded continuously over some time interval, e.g., when $T_0 = [0, 1]$.

The nature of time intervals defines whether a time series is discrete or continuous.

Q What is time series analysis?

$\Rightarrow$ Time series analysis is a quantitative forecasting method to predict future values, where numerical data are obtained at regular time intervals and projection ~~has~~ are made based on past and present observations.

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

Q Hourly temperature reading

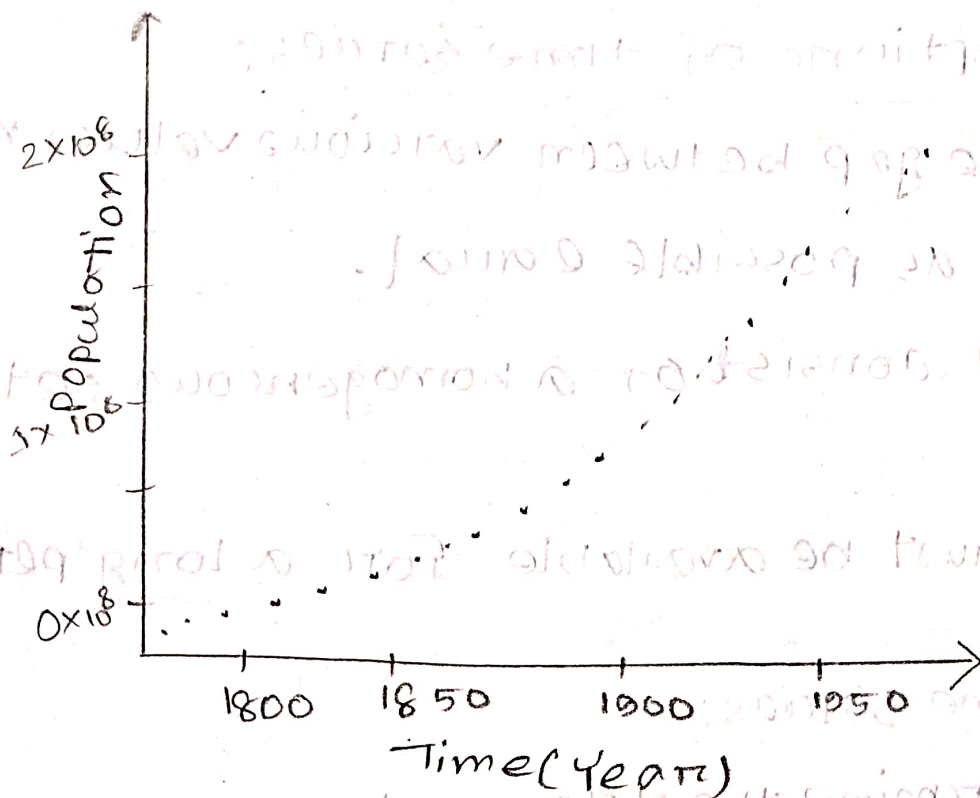
Q Daily stock prices.

Q Weekly traffic volume.

Q Annual growth rate.

Practical example: Suppose, we have the data of population of the U.S.A at 10 year interval 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 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 time. And then we can find or predict the future values of population. This is a practical example of time series analysis.



According to Ya-Ien Chou, "A time series may be defined as a collection of readings belonging to different time periods of some economic variable or composite ~~set~~ of variables."

There are two types of time series, such as -

(i) Discrete time series: A time series is said to be discrete, when observations are taken only at a specific time.

(ii) Continuous time series: A time series is said to be continuous, when observations are recorded continuously over some time interval.

Assumptions of time series:

- (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 a long period.

Uses of time series:

To determine the type and nature of the variations in data.

It Helps in planning and forecasting.

It make plans for the future.

It Helps to compare the actual current performance of accomplishments with the expected ones and analyze the causes of such variation if any.

It To prediction or estimate or forecast the behavior of the phenomenon in future which is very essential for business planning.

It To compare the changes in the values of different phenomenon at different times or places etc.

Assumptions: Time series analysis typically relies on several key assumptions, including:

(1) Stationarity: The statistical properties of the time series such as mean and variance don't change over time. This assumption is crucial for many time series models.

(2) Homoscedasticity: The variance of the residuals (the difference between observed and predicted values) is constant across all time points.

(3) Temporal Dependence: Observations at different time points are not independent and there is a systematic pattern or structure over time.

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.

Q What are the different time series components? Illustrate them graphically.
⇒ There are four time series components - they are -

(1) Trend component

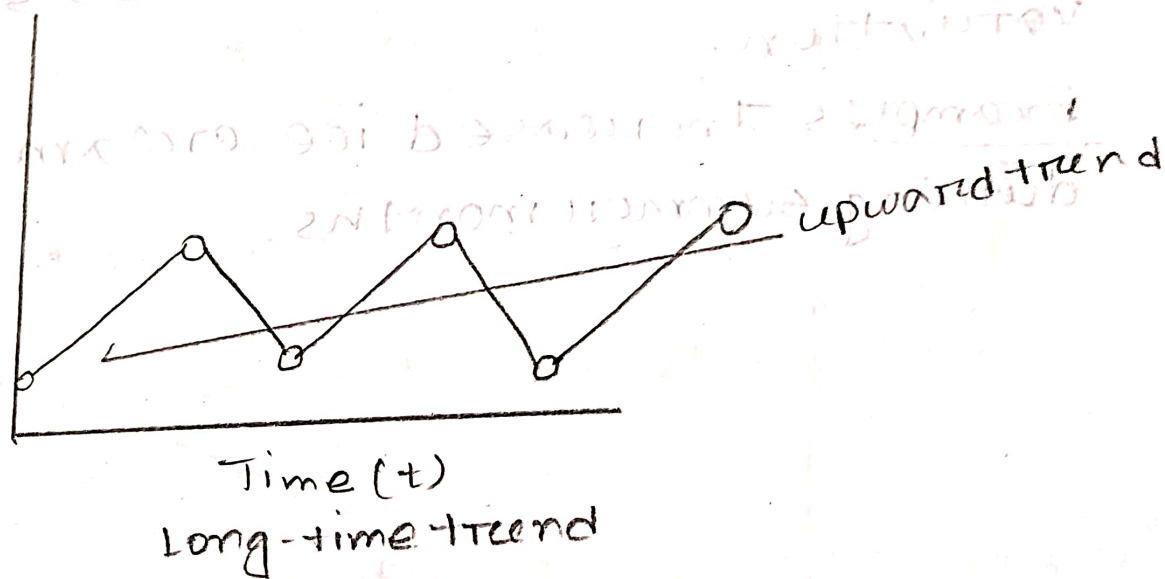
(2) Seasonal "

(3) Cyclical component

(4) Random or irregular component

Trend Component: Trend is the overall upward or 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.

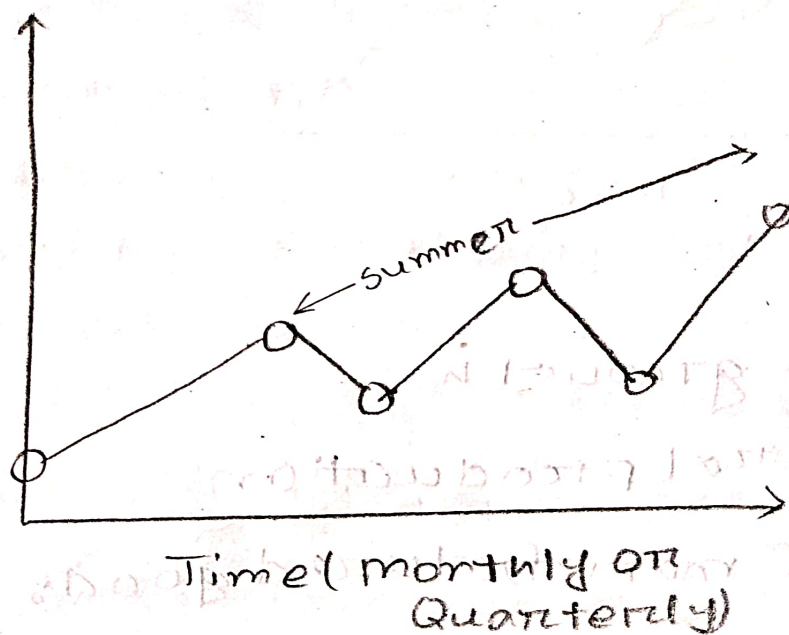
Export of manufactured goods.

Price of consumer goods.

Seasonal Component: Seasonal component is the regular patterns of upward or downward swings which is observed within 1 year. The period may be a day, a week, a quarter or a year.

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

Examples: Increased ice-cream sales during summer months.



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

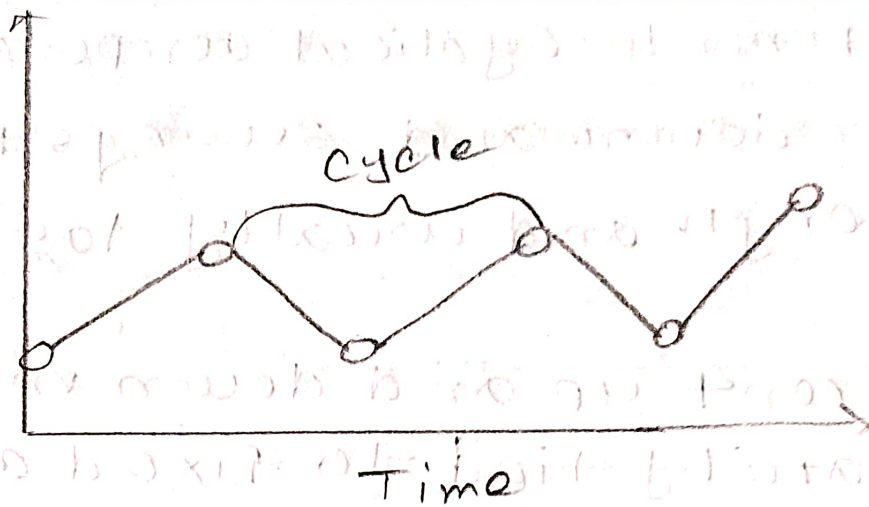
The recurrent up and down movements not necessarily tied to fixed calendar intervals. It often involves longer-term trends or economic cycles. Typically, variations are difficult to forecast because of lack of historical data. We need more than 80 years of data to make prediction about cycle.

Example: Real estate market cycles of boom and bust.

5 years of economic growth followed by 2 years of economic recession, 7 years of economic growth followed by 4 years of economic recession.

Four phases of cyclical component are -

- (1) Prosperity
- (2) Recession
- (3) Depression
- (4) Recovery.

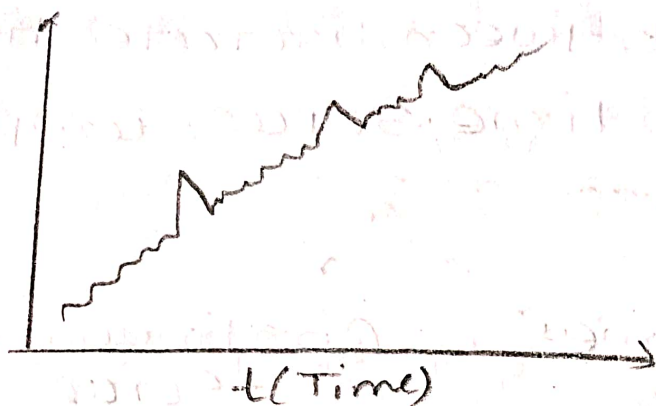


Random or irregular component: The variations which are happened due to uncertain or unusual cases (Natural disaster, accidents) called random variation. These are short term fluctuations that can't be attributed to trend, seasonality or cycles.

⇒ The irregular fluctuation can't be predicted.

⇒ Shows a sudden rise or fall at certain instants of time.

⇒ Short duration and non-repeating.



⇒ Mention different objectives of time series analysis.

- ⇒
- (1) Description of data.
 - (2) Forecasting and control.
 - (3) Hypothesis testing.
 - (4) Interpretation.
 - (5) Simulation.
 - (6) To facilitate comparison.
 - (7) To review the current achievement.
 - (8) To understand underlying forces or structure that produced the observed.

⇒ Define a time series model.

A time series model for the observed data $\{x_t, y_t\}$ is a specification of the joint distributions (or possibly only the means and covariances) of a sequence of random variables $\{x_t, y_t\}$ of which $\{x_t, y_t\}$ is postulated to be a realization.

⇒ Distinguish between discrete-time series and continuous-time series with suitable example.

⇒ Discrete-time series

1) Observations are taken only at a specified time/interval.

2) It is defined over a finite domain of sequence.

3) It is countable.

4) In which recorded observations are discrete.

5) Example: Daily stock prices or monthly sales data are discrete time series.

Continuous-time series

1) Observations are recorded continuously over some time interval.

2) It is defined over a finite or infinite domain of sequence.

3) It is uncountable.

4) In which recorded observations are continuous for a given interval.

5) Example: Temperature recorded every second or heart rate monitored continuously throughout a day.

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 far 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 those such

AR(MA)

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 characteristics of the sequence.

Second order properties of a time series model typically refer to statistical properties related to its covariance structure. These properties include autocovariance and autocorrelation functions, which describe how the values of the time series are related to their own past values at different lags. Essentially, they provide insights into the dependence structure and variability of the time series over

Q. Discuss the general approach recommended 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 ~~of~~ to the data. There are several ways in which trend and seasonality can be removed.

They are —

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

(2) Differencing the data: i.e. replacing the original series $\{x_t\}$ by $\{y_t = x_t - x_{t-1}\}$.

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 ~~inter~~inverting the transformation described above to arrive at forecasts of the original data $\{x_t\}$.

⇒ An extremely useful alternative approach is to express the series in terms of its fourier components.

□ 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.

⇒ Stationarity in a time series model means that the statistical properties of the data such as mean, variance, autocorrelation structure don't change over time. In

Simpler terms, it suggests that the data behaves consistently over time, making it easier to analyze and model. A time series has stationarity if a shift in time doesn't cause a change in the shape of the distribution. Yes, I think stationarity is an important property while conducting a time series analysis. Because -

1) Easier Analysis: Stationary time series exhibit consistent statistical properties over time, simplifying the analysis process. This allows for the use of various statistical techniques and models that assume stationarity such as autoregressive integrated moving average (ARIMA) models.

2. Meaningful Interpretation: Stationarity makes it easier to interpret trends, patterns and relationships within the data.

3. Modeling assumptions: Many time series models such as ARIMA, SARIMA and VAR assume stationarity. Violation of stationarity assumptions can lead to biased parameter estimates and unreliable forecasts.

4. Forecasting accuracy: Stationary time series are generally more predictable and stable, leading to more accurate forecasts. Most statistical forecasting methods are based on the assumption through the use of mathematical transformation.

5. Statistical tests: Stationarity is often a prerequisite for conducting various statistical tests such as unit root test which are used to determine the stationarity of a time series and guide the selection of appropriate modeling technique.

Thus, ~~ensuring~~ ensuring stationarity in time series data is essential for meaningful interpretation, reliable forecasting and so on.

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

⇒ In several of the time series data, there is a clear trend in the data but no periodic component, suggests trying a model 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. A useful technique for estimating m_t is the method of least squares. In the least square procedure, we attempt to fit a parametric family of functions e.g. -

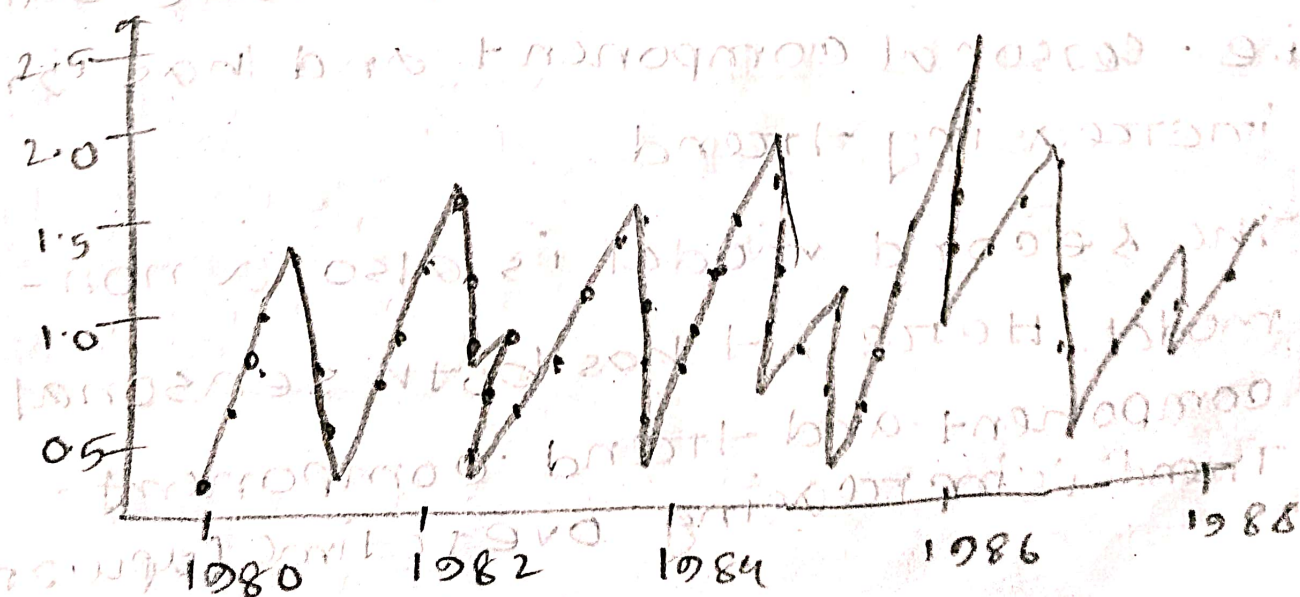
$$m_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2$$

to the data $\{x_1, \dots, x_n\}$ by choosing the parameters $(\alpha_0, \alpha_1, \alpha_2)$ to minimize $\sum_{i=1}^n (x_i - m_t)^2$. This method of curve

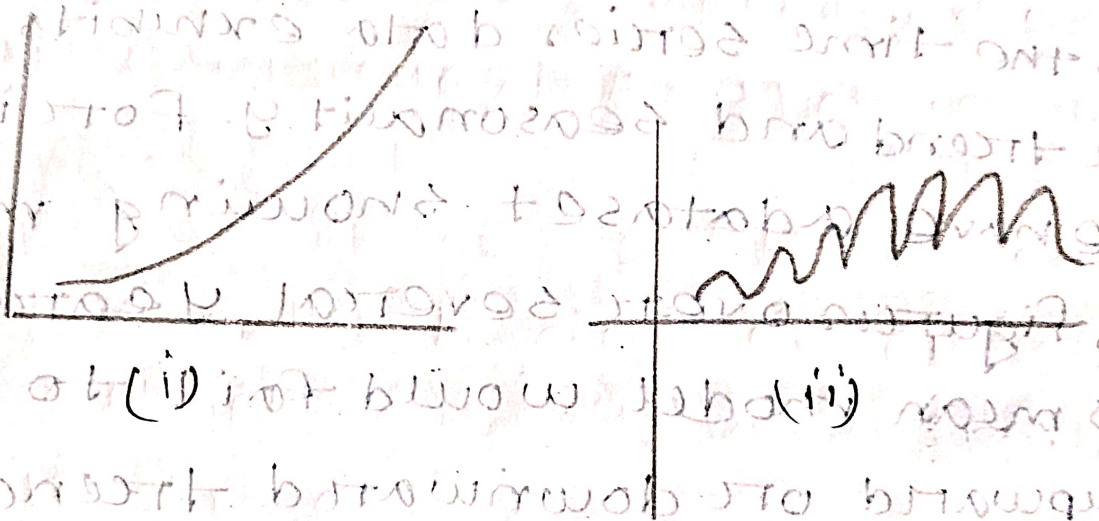
fitting is called least squares regression. It can also be used to estimate the higher order polynomial trends in the same way.

☐ When a zero-mean model is inappropriate. Explain with a suitable time series plot.

⇒ A zero mean model might be inappropriate when the time series data exhibits a clear trend and seasonality. For instance, if we have a dataset showing monthly sales figures over several years, a zero mean model would fail to capture the upward or downward trend in sales over time.



In this plot, we can see a clear upward trend in the data, indicating that the mean is not zero. Using a zero mean model in such a case would result in a poor fit to the data and inaccurate predictions.

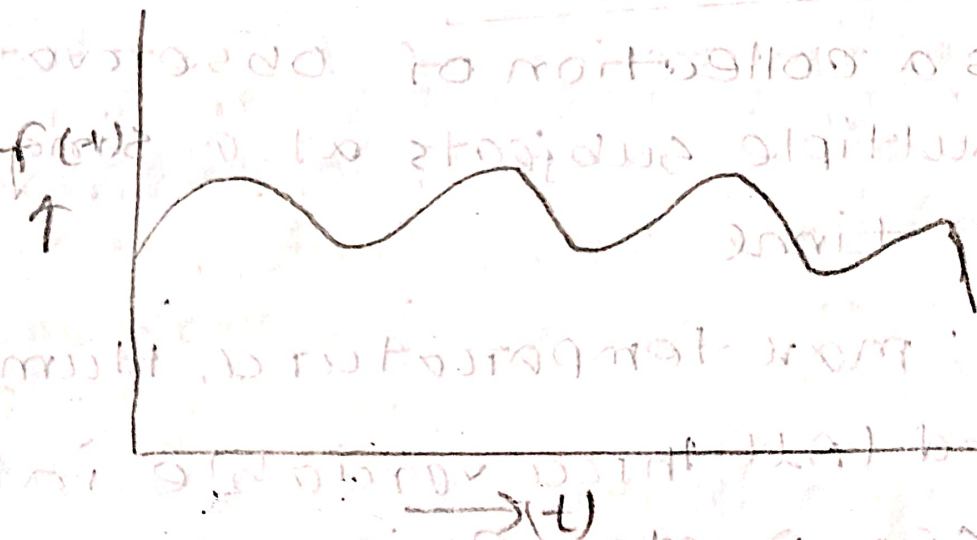


The first model is a non-stationary model. Hence, it has no periodic component i.e. - seasonal component and has an increasing trend.

The second model is also a non-stationary model. Hence, it has both seasonal component and trend component. Trend is increasing over time (upward).

Compare: The two models are non-stationary. That means we have to transform these models into stationary model. For first model, we estimate trend by method of least squares. On the other hand, for second model, we use classical decomposition which is multiplicative because the seasonal component increase with time and we also use differencing.

☐ Sketch a time series plot which exhibits stationarity.



It is a time series plot which exhibits stationarity.

☐ Explain time series, cross-sectional and panel data with example.

⇒ Time series: Time series 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 collected on first day of every year (multiple intervals of time)

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

Example: max temperature, humidity and wind (All three variable in New York, SFO, Boston, Chicago (multiple entities) on 12/05/2012 (single instance).

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

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

Q. What are the basic differences between the seasonality and cyclical variation.

Explain with examples.

Basic Differences	Seasonal variations	Cyclical variation
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 one year and is usually from 3 years to 10 or 12 years.
Causes of change	Seasonal variations are caused by changes season, traditions, holidays or cultural events	Cyclical fluctuations are caused by currency expansion or contraction, decrease or increase in sale etc.

Basic differences	Seasonal	Cyclical
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
	Seasonal variations can be measured more accurately.	Cyclical variations can't be measured accurately
Example	Sales of ice-cream increases during summer season	Macro-economic changes cause recession, unemployment

Q What is meant by deseasonalizing a time series?

⇒ Seasonal adjustment or deseasonalization is a statistical method for removing the seasonal component of time series. It is usually done when we want to analyze the trend and cyclical deviations from trend.

of a time series independently of the seasonal components. It helps in identifying long-term changes or irregular fluctuations in the data that are not related to seasonal variations.

▣ What are the different techniques to estimate and eliminate the trend component in the absence of seasonality?

⇒ 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 squares.

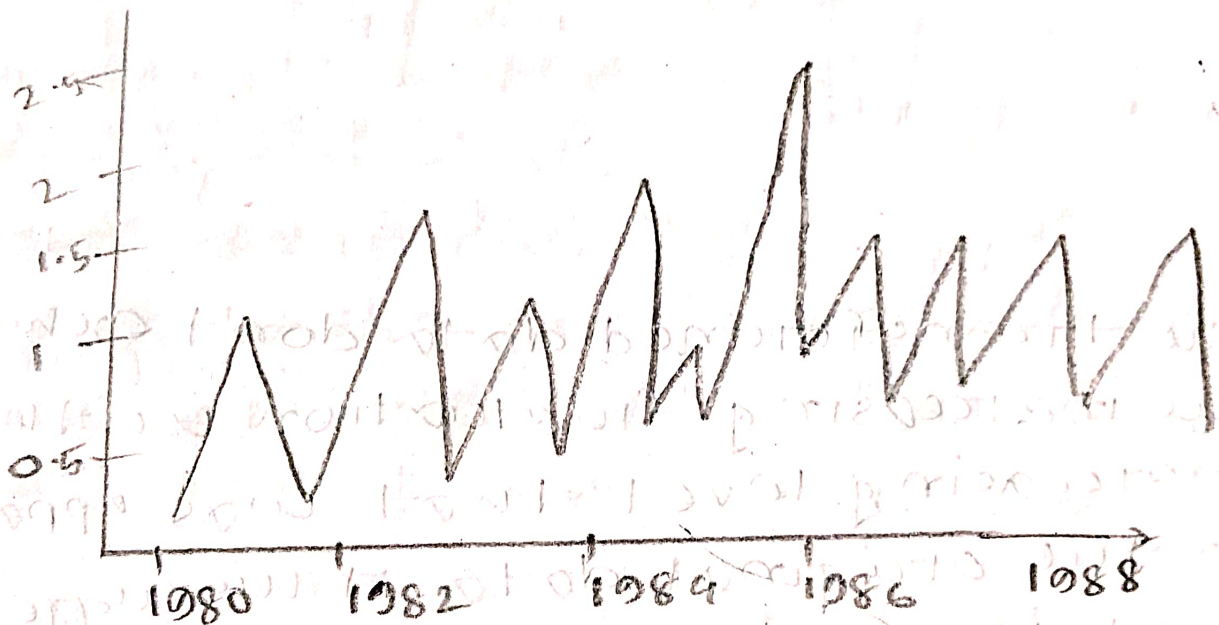
III If the estimated residuals have no trend and show smoothness, how would you quantify such feature?

→ 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 the future values of the series. We quantify such dependence by the autocorrelation function as a measure of dependence and stationary processes as a family of useful models exhibiting a wide variety of dependence structure.

III With a suitable example, explain the concept of preliminary transformation for a given time series data. When it is required?

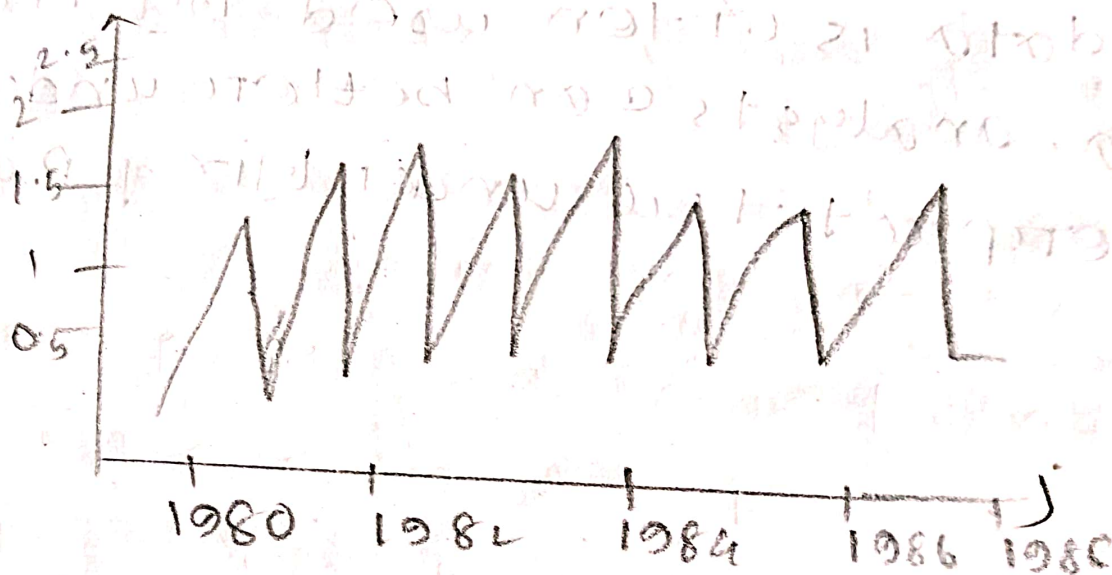
→ Preliminary transformation of a time series involves applying

mathematical operations to the data to improve its properties for analysis. 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. By transforming the data, analysts can better understand and interpret the underlying patterns.



This figure shows the monthly sales of red wine by Australian winemakers from January 1980 through October 1988. In this data, there is upward trend and seasonal variation. Seasonal and noise fluctuations appear to increase ~~the~~

with the level of the process. After doing preliminary transformation of wine data by applying log transformation we will get the transformed data.



The transformed data don't 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 data than for the original series.

Q What is STL Decomposition?

⇒ STL (Seasonal and Trend decomposition using LOESS) is a non-parametric generalisation of ordinary least squares regression. The fitting technique doesn't require a prior specification of the relationship between the dependent and independent variables. The STL works by iterating through the smoothing the trend and seasonal components. STL is a procedure for regular time series so that the design points of the smoothing operation are equally spaced.

Q What are the advantages of using STL decomposition over the classical SEATS and X-11 methods of decomposition?

⇒ flexibility: STL is more flexible and robust in handling various types of data including non-linear trends and irregular seasonal patterns.

2. No assumption of stationarity: Unlike the classical methods, STL doesn't assume that the data to be stationary, making it suitable for analyzing time series with ^{changing} dynamics.

3. Transparency: STL provides a clear separation of trend, seasonal and random components, making it easier to interpret.

4. Robustness to outliers: STL is less sensitive to outliers compared to classical methods as it uses a robust smoothing technique to estimate the trend component.

STL decomposition is a versatile and powerful method for time series analysis, especially when dealing with complex and non-standard patterns in the data.

Q: What is white noise and IID noise?

⇒ white noise: Refers to a series of random variables that are uncorrelated with each other and have constant variance. Essentially, each error term is independently and identically distributed with no pattern or structure over time. It's like flipping a fair coin repeatedly, each flip is independent of the others.

IID noise: Stands for independent and identically distributed noise, indicating that each error term in the series is drawn from the same probability distribution and is independent of the others. While IID noise shares characteristics with white noise, it doesn't necessarily mean that the noise is uncorrelated over time, as it could still exhibit patterns or

white noise: sample could follow different probability distribution.

Usually used to describe the behaviour of an error term in a model.

for example, when modeling stationary time series:

$$x_t = \alpha + \theta e_{t-1} + e_t$$

e_t follows a white noise process if

$$\text{Cov}[e_t, e_{t-j}] = 0 \quad \forall j \neq 0$$

IID noise: special case of white noise. each

sample has the same probability distribution

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

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

Q] What is autocovariance function of a given time series?

⇒ The autocovariance function in time series analysis measures the covariance between observations of a series at different time points. It quantifies how the observations at one time point relate to observations at other time points, capturing both the direction and strength of this relationship.

Mathematically, for a time series $\{x_t\}$, the autocovariance function $\gamma(h)$ at lag h is defined as:

$$\gamma(h) = \text{Cov}(x_t, x_{t+h})$$

Autocovariance function measures autocorrelation among error term in

$$\gamma_x(h) = \gamma_x(h, 0) = \gamma_x(t+h, t)$$

$\gamma_x(\cdot) \rightarrow$ autocovariance function

$\gamma_x(h) \rightarrow$ value of $\gamma_x(\cdot)$ at lag h .

Q What is autocorrelation function?

⇒ Autocorrelation function measures the correlation between the series and its lagged values, rather than their ~~series~~ covariance. The autocorrelation function is normalized, which makes it easier to interpret compared to autocovariance function. The autocorrelation function is derived from the autocovariance function. Let $\{X_t\}$ be a stationary time series. Its mean is $\mu_X = \mu_X(t)$. Its autocovariance function of $\{X_t\}$ at lag h is -

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

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

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

A sample ACF that is close to zero for all non-zero lags suggests that an appropriate model for the data might be iid noise.

Q How does autocovariance function differ from the autocorrelation function?

$\Rightarrow$ The autocovariance function measures the linear dependence between two observations at different time points without standardization. It gives the raw measure of dependence showing whether the series tends to increase or decrease together at different lags. It's also a measure of how the covariance between observations changes as the lag increases.

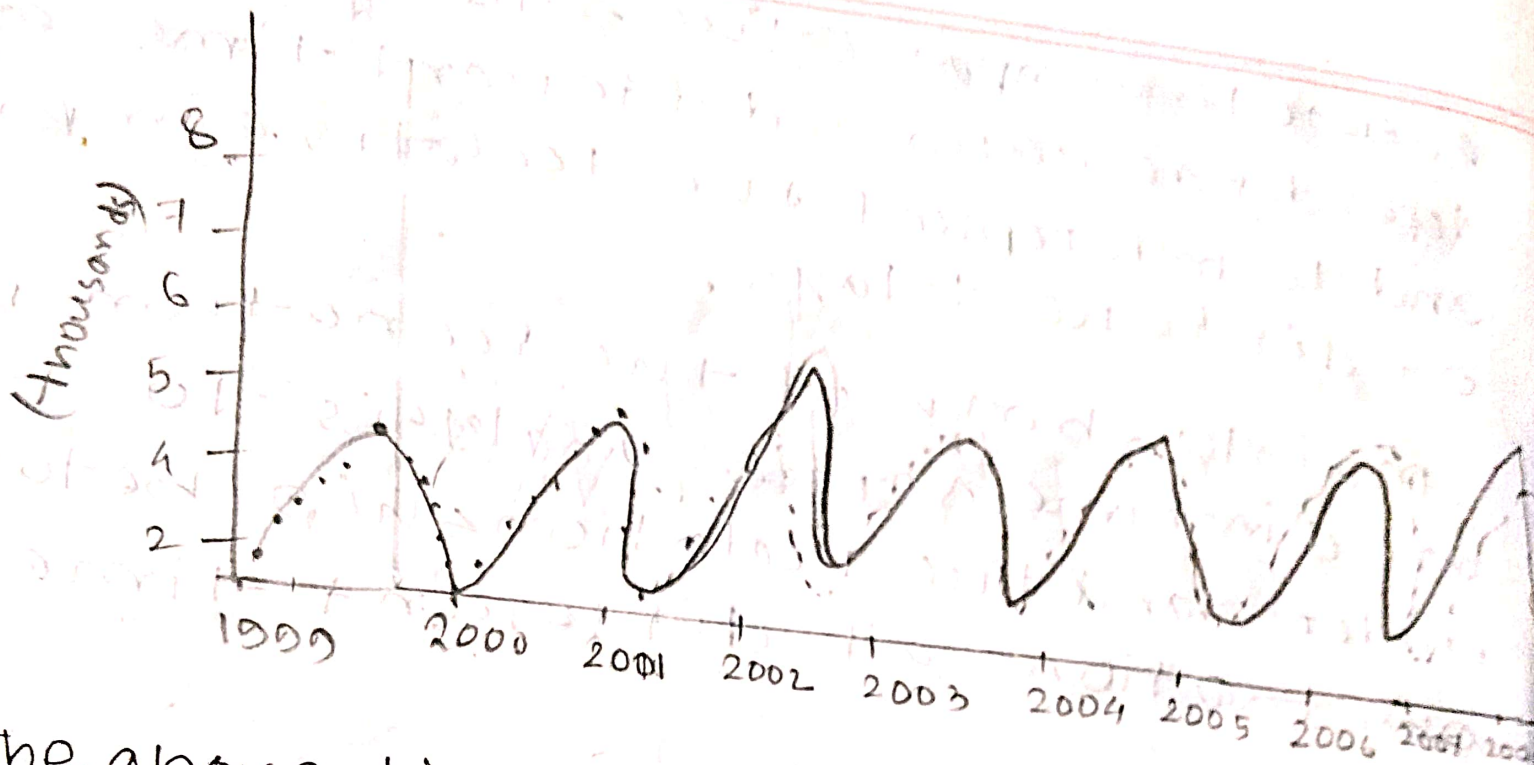
On the other hand, the autocorrelation function is derived from the autocovariance function by normalizing it with the variance of time series. It provides a standardized measure of the strength and direction of dependence at different lags, making it

easier to compare the strength of dependence across different time series and to interpret the relative importance of different lags.

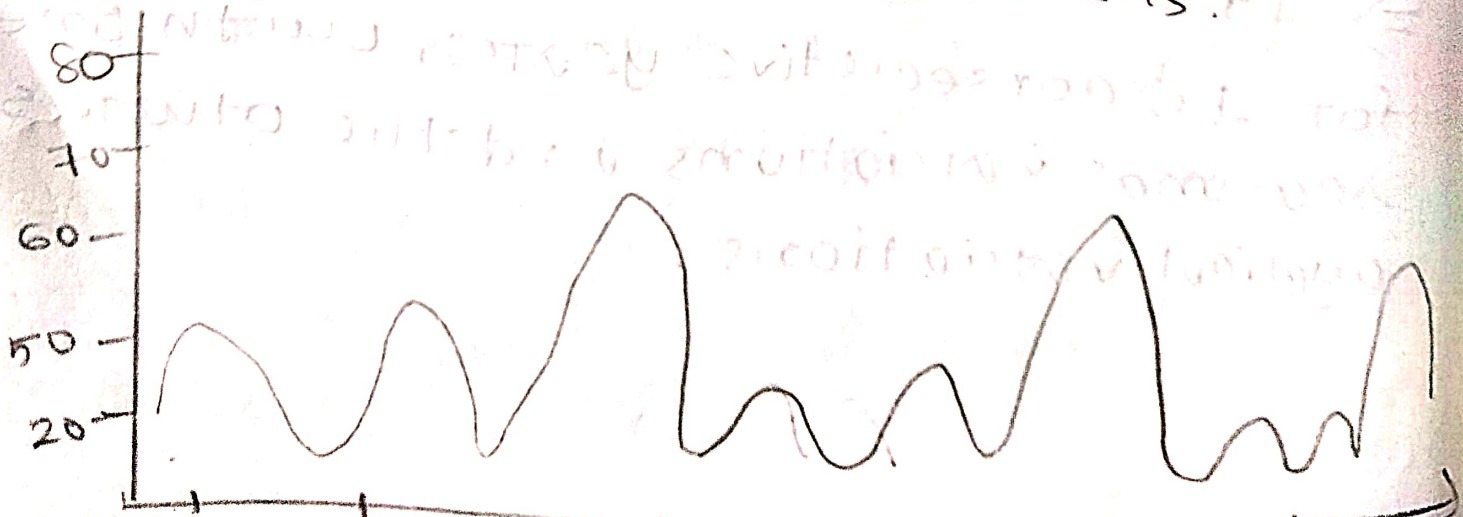
mainly, both of the function is used in time series analysis to understand the relationship between observations at different time points.

□ Suppose, you have two time series data for the 10 consecutive years. One has the seasonal and the other has the cyclical variations. Show the pattern of the time series plots in a graph with proper interpretations.

⇒ If we have two time series data for 10 consecutive years, with one showing seasonal variations and the other showing cyclical variations.



The above time series plot is exhibiting patterns that repeat within each year such as quarterly or monthly fluctuations due to seasonal factors like weather, holidays or other regular events. For example, sales of ice-cream might increase in the summer months and decrease in the winter months.



The above graph shows longer-term patterns of expansion and contraction, typically spanning multiple years, representing ~~multiple years~~ economic cycles. Over a period of several years, stock prices may rise and fall in cycles, reflecting broader economic conditions. During an economic expansion, stock prices tend to rise as companies experience growth and profitability increases. Conversely, during an economic recession, stock prices may decline as economic activity slows down and companies struggle. These cycles can last for several years.

☐ Explain different types of decomposition, by mentioning the suitability of each with example.

⇒ There are a lot of different decomposition algorithms, here we explain a very simple but useful approach. The

general mathematical model for this decomposition is -

$$Y_t = f(S_t, T_t, E_t)$$

where, S_t is the seasonal component, T_t is the trend effect (sometimes called the trend-cycle effect) and E_t is the random error component. There are usually two forms for the function f ; an additive model -

$$Y_t = S_t + T_t + E_t$$

and a multiplicative model -

$$Y_t = S_t \cdot T_t \cdot E_t$$

The additive model is appropriate if the magnitude (amplitude) of the seasonal variation doesn't vary with the level of the series, while the multiplicative version is more appropriate if the amplitude of the seasonal fluctuation increases or decreases with the average level of the time series.

Decomposition is useful for breaking a time series down into component parts. For the additive model, it is relatively easy. First, we would model and remove the trend. A simple linear model could be used to do this. Say, $T_t = \beta_0 + \beta_1 t$. Other methods could also be used - moving averages can be used to isolate a trend and remove it from the original data as could more sophisticated regression models. These techniques might be appropriate when the trend is not a straight line over the history of the time series. Differencing could also be used, although it is not typically in the classical decomposition approach. Once the trend or trend-cycle component is estimated, the series is detrended -

$$Y_t = T_t + E_t$$

There are also some decomposition methods. Such as—

SEATS decomposition (seasonal extraction in ARIMA time series)

This method is used for time series decomposition specifically designed for seasonal adjustment. It's often employed by statistical agencies to remove seasonal effects from economic data such as GDP, employment figures or retail sales. This method works only with quarterly and monthly data. Seasonality of other kinds, such as daily data or hourly data or weekly data requires an alternative approach.

STL decomposition: Local Regression method or Seasonal Decomposition → a non-parametric generalisation of ordinary least squares regression. The fitting technique doesn't require a

prior specification of the relationship between the dependent and independent variables. The seasonal and trend decomposition using the LOESS method abbreviated as STL works by iterating through the smoothing of the seasonal and trend components. STL is a procedure for regular time series so that the design points of the smoothing operation are equally spaced.

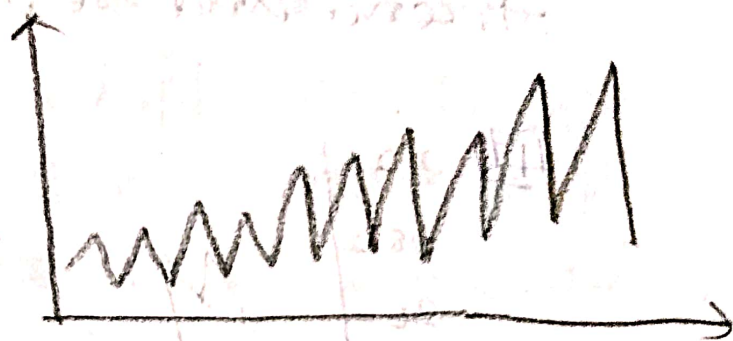
X-11 Decomposition: This method works only quarterly and monthly data. It is based on classical decomposition but includes many extra steps and features in order to overcome the drawbacks of classical decomposition. It has some sophisticated methods for handling trading day variation, holiday effects and the effects of known predictors. It handles both additive and multiplicative decomposition. The process is entirely automatic and tends to be highly

robust to outliers and level shifts in the time series.

☐ Why a stochastic process provides a formal way to examine time series variables?

⇒ Stochastic processes offer a structured framework to model random variations in time series data, making it easier to analyze and predict future behaviour. By incorporating randomness into the model, it accounts for uncertainty and captures the inherent volatility and unpredictability often observed in real-world data. This formal approach allows for statistical inference and hypothesis testing, providing insights into the underlying dynamics of the time series variables.

tion:

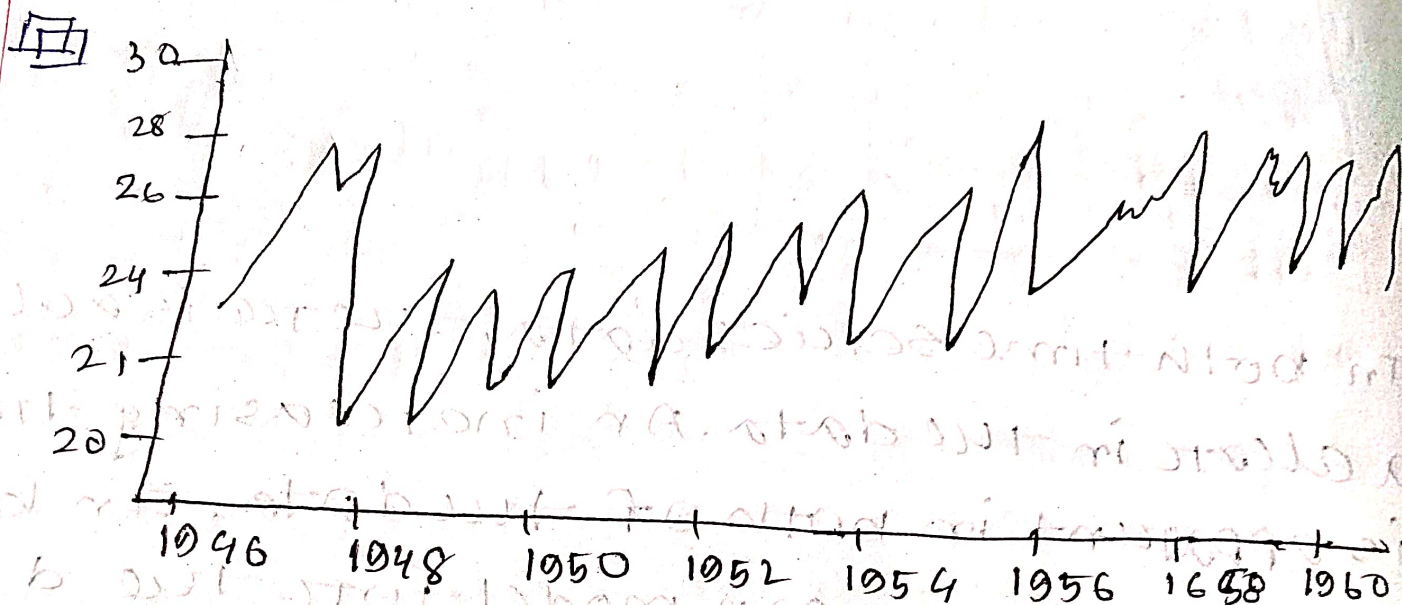


both time series data, there is a clear trend in the data. An increasing trend is apparent in both of the data. In both cases, a zero-mean model for the data is clearly inappropriate. In the first data, there is no apparent periodic component, suggests trying a model of the form.

$$X_t = m_t + Y_t$$

where m_t is a slowly changing function, known as the trend component and Y_t has zero mean. A useful technique for estimating m_t is the method of least squares.

In the second graph, there is both trend and seasonal component.



Describe the following figure.

⇒ From the graph, we can say that the time series data is a non-stationary data where statistical properties like mean, variance is not constant over time. As the variance is increasing over time, then multiplicative decomposition will be suitable for deseasonalizing the data. We can also use differencing to remove trend and seasonality from the above time series data.

LAKE.DAT Data:

A graph of the level in feet of Lake Huron in the years 1875-1972 is displayed first. Since the lake level appears to decline at a roughly linear rate, ITSM was used to fit a model of the form -

$$X_t = \alpha_0 + \alpha_1 t + Y_t : t = 1, \dots, 98$$

The least squares estimates of the parameter values are -

$$\hat{\alpha}_0 = 10.202 \text{ and } \hat{\alpha}_1 = -0.0242$$

The estimates of the noise, Y_t in the model are the residuals obtained by subtracting the least squares line from X_t . 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. In particular, there are long stretches of residuals that have the same sign. This would be very unlikely to occur if the residuals were observations of iid noise (with mean zero). Smoothness of the

A group 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. If we were to assume the validity of the fitted model with iid residuals $\{y_t\}$, then the minimum mean squared error predictor of the next residual would be zero.

Seasonal

Variations got some degree of regularity.

Involves short-term period

Almost identical or same pattern arises every year.

(4) Helps to predict demand of the product as per upcoming time.

Cyclical

(1) No regular variations are recorded.

(2) Involves longer duration

(3) Patterns completely depend upon market conditions.

(4) Helps to formulate policy as per market

Components of time series:

Trend: The general tendency of the time series data to increase or decrease during a long period of time is called the secular trend or long-term trend or simply trend. The concepts of trend doesn't include short range oscillations. This is true in most of business and economic statistics. Trend may have a downward or upward movement such as production prices, income etc. upward trend while a downward trend is noticed in the time series relating to deaths, ~~pedid~~ epidemics etc. Trend is the general, smooth, long term.

There are different types of trend -
Linear or straight line trend: If we get a straight line, when the values are plotted on a graph, then it is called a linear ~~mode~~ trend.

Non-linear trend: If we get a curve after plotting the time series values then it is called non-linear.

Seasonal variation: A variation which occurs with some degree of regularity within a specified period of one year or less than one year is called seasonal variation. There are hourly, daily, monthly or half yearly variations also.

The seasonal variations may be attributed to the following two causes:

- (1) Those resulting from natural forces i.e. weather and seasons.
- (2) Those resulting from non-made conventions, i.e. Eid day, Durga puja.

Cyclical variation: The oscillatory movement in a time series with period of oscillation more than one year are termed as cyclic variations. One complete period is called a cycle.

Random or irregular variation: Apart from the regular variations, almost all the series contain another factor called the random or irregular or residual variations which are not accounted for secular trend, seasonal and cyclic variations. These variations are happened due to uncertain and unusual cases. It is purely random and beyond the control of human hand but a part of system. These are caused by earthquakes, water, floods etc.

Q. What are the different types of model used in time series?

⇒ The following are the models commonly used for the decomposition of a time series into its components.

- (i) Decomposition by additive hypothesis.
- (ii) Decomposition by multiplicative hypothesis.
- (iii) mixed models.

Decomposition by Additive hypothesis:

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:

- (1) S_t , C_t , U_t operate with absolute

Sta effect irrespective of the trend.

(i) C_t (and S_t) will have positive or negative values and the total of positive and negative values for the cycle (and any year) will be zero.

(ii) U_t will also have positive or negative values in the long run $\sum U_t$ will be zero.

Decomposition by multiplicative hypothesis;

According to the multiplication 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 the long term period are unity.

Mixed model: The different combination of additive and multiplicative models are named as mixed model defined as -

$$Y_t = T_t C_t + S_t U_t$$

$$Y_t = T_t + S_t C_t U_t$$

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

Time plot: The first step in analyzing a time series is to plot the observations against time. This plot is known as time plot. It is similar plot of scatter diagram. This will show up important features such as trend, seasonality, discontinuities and outliers etc.

Time series model: A time series model for the observed data $\{y_t\}$ is a specification of the joint distributions (or possibly only the means and covariances) of a sequence of random variables $\{X_t\}$ of which $\{y_t\}$ is postulated to be a realization.

Some zero mean model:

IID noise: The simplest model for a time series is one in which there is no trend or seasonal component and in which the observations are simply independent and identically distributed (iid) random variables with mean zero. We refer to such a sequence of random variables $x_1, \dots, x_n$ as iid noise. By definition, we can write, for any positive integer n and real numbers $y_1, \dots, y_n$

$$\begin{aligned} P[Y_1 \leq y_1, Y_2 \leq y_2, \dots, Y_n \leq y_n] &= P[Y_1 \leq y_1] \cdot P[Y_2 \leq y_2] \cdot \dots \cdot P[Y_n \leq y_n] \\ &= F(y_1) \cdot F(y_2) \cdot \dots \cdot F(y_n) \end{aligned}$$

where $F(\cdot)$ is the cumulative distribution function of each of the identically distributed random variables $Y_1, Y_2, \dots, Y_n$. In this model, there is no dependence between observations. In particular, for all $n \geq 1$ and all $y_1, y_2, \dots, y_n$

$$\begin{aligned} P[Y_{n+1} \leq y \mid Y_1 \leq y_1, Y_2 \leq y_2, \dots, Y_n \leq y_n] &= \\ P[Y_{n+1} \leq y] \end{aligned}$$

It shows that knowledge of $Y_1, Y_2, \dots, Y_n$ is of no value for predicting the behaviour of Y_{n+1} . Given the values of $Y_1, Y_2, \dots, Y_n$ the function f that minimizes the mean squared error,

$E[\{Y_{n+1} - f(Y_1, Y_2, \dots, Y_n)\}^2]$ is in fact identically zero.

Although this means that IID noise is a rather uninteresting process for forecasters, it plays an important role as a building block for more complicated time series models.

Example (Binary process): Let us consider a sequence of iid random variables $\{Y_t, t=1, 2, \dots\}$ with $P\{Y_t = 1\} = p$ and $P\{Y_t = -1\} = 1 - p$ where $p = \frac{1}{2}$. The time series obtained by tossing a penny repeatedly and scoring +1 for each head and -1 for each tail is

usually modeled as a realization of this process.

Random walk: The random walk $\{S_t, t = 0, 1, 2, \dots\}$ is obtained by cumulatively summing (or integrating) iid random variables. Thus a random walk with zero mean is obtained by defining $S_0 = 0$ and

$$S_t = Y_1 + Y_2 + \dots + Y_t; t = 1, 2, \dots$$

where $\{Y_t\}$ is IID noise. If $\{Y_t\}$ is the binary process then $\{S_t, t = 0, 1, 2, \dots\}$ is called a simple symmetric random walk.

models with trend and seasonality: The

graph of the population data, which contains no apparent periodic component, the model will be of the form -

$$Y_t = T_t + U_t$$

where T_t is a slowly changing function known as the trend component and U_t has zero mean. T_t is estimated by least square methods.

Weakly Stationary: A time series $\{Y_t, t = 0, \pm 1, \pm 2, \dots\}$ is said to be weakly stationary if its mean function and covariance function is independent of time t , i. e. -

- (i) $\mu_Y(t)$ is independent of time t and
- (ii) $\gamma_Y(t+h, t)$ is independent of time t for each h .

Usually weakly stationary is termed as simply stationary.

Strictly Stationary: A time series $\{Y_t, t = 0, \pm 1, \pm 2, \dots\}$ is said to be strictly stationary if the joint distribution of $Y_1, Y_2, \dots, Y_n$ is the same as the joint distribution of $Y_{1+h}, Y_{2+h}, \dots, Y_{n+h}$ for all integers h and $n > 0$, i. e. -

$$\{Y_1, Y_2, \dots, Y_n\}^T \stackrel{d}{=} \{Y_{1+h}, Y_{2+h}, \dots, Y_{n+h}\}^T$$

mean function of Time series;

Let $\{Y_t\}$ be a time series with $E\{Y_t^2\} < \infty$, then the mean function of $\{Y_t\}$ is -

$$\mu_Y(t) = E\{Y_t\}.$$

Co-variance-function: The co-variance

function of $\{Y_t\}$ is defined as -

$$\gamma_Y(\pi, s) = \text{cov}(Y_\pi, Y_s) = E\{[Y_\pi - \mu_Y(\pi)][Y_s - \mu_Y(s)]\}$$

for all integers π, s and t .

Properties of strictly stationary time series $\{Y_t\}$:

The main properties of $\{Y_t\}$ are given below -

(i) The random variables Y_t are identically distributed.

(ii) $\{Y_t, Y_{t+n}\}^T \stackrel{d}{=} \{Y_1, Y_{1+n}\}^T$; for all integer t and n .

(iii) $\{Y_t\}$ is weakly stationary if $E\{Y_t^2} < \infty$ for all t .

(iv) weak stationary doesn't imply strict stationary.

(v) An iid sequence is strictly stationary.

Show that a strictly stationary process with $E\|Y_t\|^2 < \infty$ is weakly stationary or properties of weakly stationary time series.

Proof:

If $E\|Y_t\|^2 < \infty$ then the properties of strictly stationary are—
i) the random variables Y_t are identically distributed

ii) $\|Y_t, Y_{t+h}\|^T \stackrel{d}{=} \|Y_1, Y_{1+h}\|^T$ for all integers t and h

we can write $E\|Y_t\|^2$ is independent of t and $\text{Cov}\|Y_t, Y_{t+h}\| = \text{Cov}\|Y_1, Y_{1+h}\|$ which is also independent of t .
thus, $\{Y_t\}$ is weakly stationary.

Auto-covariance function: Let $\{Y_t\}$ be

any stationary time series. Then auto-covariance function (ACVF) of $\{Y_t\}$ is defined as

$$\gamma_Y(h) = \text{Cov}(Y_{t+h}, Y_t)$$

$$= E[(Y_{t+h} - E(Y_{t+h}))(Y_t - E(Y_t))]$$

Auto correlation function: Let $\{X_t\}$ be any stationary time series. Then the auto-correlation function of $\{X_t\}$ is defined as -

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

Linearity property of Co-variance:

If $E\{X^2\} < \infty$, $E\{Y^2\} < \infty$, $E\{Z^2\} < \infty$ and a, b and c are any real constants, then

$$\text{Cov}(aX + bY + c, Z) = a\text{Cov}(X, Z) + b\text{Cov}(Y, Z) + 0$$

Since,

$$\begin{aligned} \text{Cov}(c, Z) &= E[(Z - E(Z))(c - E(c))] \\ &= E[(Z - E(Z)) \cdot c - c^2] \\ &= E[(Z - E(Z)) \cdot c] - E[c^2] \end{aligned}$$

$$= 0$$

Example: If $\{X_t\}$ is IID noise and $E\{X_t^2\} = \sigma^2 < \infty$, then it is obvious that $E\{X_t\} = 0$ for all t , so it satisfies the first requirement of weakly

Stationary:

By the assumption $\{x_t\}$ are independent,

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

which doesn't depend on t ,

Here IID noise with finite second moment is stationary, i.e.

$\{x_t\} \sim \text{IID}(0, \sigma^2)$. It indicates that the random variables $\{x_t\}$ are independent and identically distributed random variables with mean 0 and variance σ^2 .

Show that, $\gamma_r(t+h, t) = \sigma^2$ when $h=0$,

Proof: By definition, we have -

$$\gamma_r(t+h, t) = \text{cov}(x_{t+h}, x_t)$$

$$= E[(x_{t+h} - E(x_{t+h})) (x_t - E(x_t))]$$

$$= E[(x_{t+h} - E(x_{t+h})) (x_t - E(x_t))];$$

$$= E[x_{t+h} x_t] = \sigma^2 \quad \text{when } h=0$$

Since, $E(x_t) = 0$

[Proved]

White noise: If $\{Y_t\}$ is a sequence of uncorrelated random variables, each with zero mean and variance σ^2 , then clearly $\{Y_t\}$ is stationary with the same covariance function as the IID noise as:

$$\gamma_Y(h) = \begin{cases} \sigma^2; & \text{if } h=0 \\ 0; & \text{if } h \neq 0 \end{cases}$$

which doesn't depend on t . Such a sequence is referred to as white noise. This is indicated by the notation -

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

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

Random walk: The random walk $\{S_t; t=0,1,2,\dots\}$ (starting at zero) is obtained by cumulatively summing or integrating iid random variables -

$$S_t = Y_1 + Y_2 + \dots + Y_t$$

Let $\{S_t\}$ is the random walk and $\{Y_t\}$ is iid noise with $E\{Y_t^2\} = \sigma^2 < \infty$ for all t .

Obviously, $E[Y_t] = 0$ and $\text{Cov}(Y_t, Y_t) = \sigma^2 < \infty$
 we have,

$$\begin{aligned} E[S_t] &= E[Y_1 + Y_2 + \dots + Y_t] \\ &= E[Y_1] + E[Y_2] + \dots + E[Y_t] \\ &= 0 \end{aligned}$$

$$\text{and, } \gamma_s(t+h, t) = \text{Cov}(S_{t+h}, S_t)$$

$$\begin{aligned} &= \text{Cov}(S_t + Y_{t+1} + \dots + Y_{t+h}, S_t) \\ &= \text{Cov}(S_t, S_t) \quad \text{since, } \text{Cov}(S_{t+j}, S_t) = 0 \quad ; j \geq 1 \\ &= E(S_t^2) \quad \text{since } E(S_t) = 0 \\ &= E(Y_1 + Y_2 + Y_3 + \dots + Y_t, Y_1 + Y_2 + \dots + Y_t) \\ &= E(Y_1^2) + E(Y_2^2) + \dots + E(Y_t^2) + \text{all cross terms} \\ &= E[Y_1^2] + E[Y_2^2] + \dots + E[Y_t^2] + \text{all cross terms} \\ &= \sigma^2 + \sigma^2 + \dots + \sigma^2 \quad \text{since all cross terms being zero} \\ &= t\sigma^2 \end{aligned}$$

So, $\gamma_s(t+h, t)$ depends on t . So, that
 the random walk $\{S_t\}$ is non-stationary.

Exponential Smoothing: Exponential smoothing was proposed in the late 1950s (Brown, 1959; Holt, 1957; Winters, 1960) and has motivated some of the most successful forecasting methods. It is a time series forecasting approach. It is used for univariate time series data. Forecasts produced using exponential smoothing methods are weighted averages of past observations, with the weights decaying exponentially as the observations get older. In other words, the more recent observations, the higher the associated weight.

For short-term projections, exponential smoothing is a reasonably reliable forecasting strategy. The approach gives more weight to recent data while lowering weights exponentially as the observations go further apart.

Long-term forecasts produced by this approach are slightly unreliable. Exponential smoothing's methods are appropriate for non-stationary data (i.e. data with a trend and seasonal data).

⇒ most effective when the time series parameters vary slowly over time.

Single Exponential Smoothing: used with univariate data with no trend and no seasonal pattern. Needs a single parameter called alpha (α), also known as the smoothing factor.

$0 < \alpha < 1$.

α controls the rate at which the influence of past observations decreases exponentially.

Double Exponential Smoothing:

- ⇒ Used when the data has a linear trend but no seasonal pattern. Two smoothing factors: α and β , $0 < \alpha < 1$ and $0 < \beta < 1$
- ⇒ β controls the decay of the influence of change in trend.
- ⇒ The method supports trends that change in additive ways (smoothing with linear trend) and trends that change in multiplicative ways (smoothing with exponential trend).

Triple Exponential Smoothing: Most advanced and is used when the data has linear trend and seasonal patterns. The technique applies exponential smoothing three times - level smoothing, trend smoothing and seasonal smoothing.

Three smoothing factors: α , β and γ .
 $0 < \alpha < 1$, $0 < \beta < 1$ and $0 < \gamma < 1$.
 γ controls the influence of seasonal component.

Also, called Holt-Winter's exponential smoothing, named after its contributors Charles Holt and Peter Winter.

Holt-Winter's Exponential smoothing has two categories depending on the nature of the seasonal component

(a) Holt-Winter's Additive method (for seasonality that is additive)

(b) Holt-Winter's multiplicative method (for seasonality that is multiplicative)

How to Configure Exponential smoothing:

⇒ Need to specify all the model hyperparameters explicitly. This can be challenging for both beginners and experts. Numerical optimization is commonly used to search for and find the smoothing factors (α , β and γ) for the model resulting in the most negligible error.

⇒ An exponential smoothing method can obtain values for unknown parameters by estimating them

from the observed data. The initial values and unknown parameters can be estimated by minimizing the sum of the squared errors (SSE)

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

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

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

specifying the value of α . Application of (1) and (2) is often referred to as exponential smoothing, since the recursions imply that for $t \geq 2$, $\hat{m}_t = \sum_{j=0}^{t-2} \alpha(1-\alpha)^j x_{t-j} + (1-\alpha)^{t-1} x_1$, a weighted moving average of $x_t, x_{t-1}, \dots$ with ~~the~~ weights decreasing exponentially (except for the last one)

Benefits of Exponential Smoothing: The

moving average method weights all past observations equally when they fall within the moving average window and

it gives observations outside the window zero weight.

▣ Estimation and elimination of both trend and seasonality.

(I) Classical Decomposition :- in which trend and/or seasonality are estimated and subtracted from the data to generate a noise sequence.

Explanation:

Classical decomposition model -

$$X_t = m_t + S_t + Y_t, \quad t = 1, 2, \dots, n$$

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

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

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 $d = 2\alpha$, then we ~~have~~ use -

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

where, $\alpha \leq t \leq n - \alpha$

If the period d is odd, say $d = 2\alpha + 1$, then we use (simple moving avg.)

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

The second step is to estimate the seasonal component

For each $k = 1, 2, \dots, d$, we compute the average w_k of the deviations

$$\{ (x_{k+jd} - \hat{m}_{k+jd}), \alpha \leq k+jd \leq n - \alpha \}$$

Since, these average deviations don't necessarily sum to zero, we estimate the seasonal component $\hat{s}_k$ as -

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

$$\text{and } \hat{s}_k = \hat{s}_{k-d}, k > d$$

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

$$d_t = x_t - \hat{s}_t ; t = 1, 2, \dots, n$$

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 components the estimated noise series is then given by -

$$\hat{r}_t = x_t - \hat{m}_t - \hat{s}_t ; t = 1, 2, \dots, n$$

The reestimation of the trend can be done in order to have a parametric form of the trend that can be extrapolated for the purpose of prediction and simulation.

2nd Tutorial

Time series model: A time series model for the observed data $\{Y_t\}$ is a specification of the joint distributions (or possibly only the means and co-variance) of a sequence of random variables $\{Y_t\}$ of which $\{Y_t\}$ is postulated to be a realization.

IID Noise: The simplest model for a time series is one in which there is no trend or seasonal component and in which the observations are simply independent and identically distributed (iid) random variables with mean zero. We refer to such a sequence of random variables $Y_1, Y_2, \dots, Y_n$ as IID noise.

By definition we can write, for any positive integer n and real numbers $y_1, y_2, \dots, y_n$.

$$P[Y_1 \leq y_1, Y_2 \leq y_2, \dots, Y_n \leq y_n] = P[Y_1 \leq y_1] \cdot P[Y_2 \leq y_2] \cdot \dots \cdot P[Y_n \leq y_n]$$
$$= F(y_1) \cdot F(y_2) \cdot \dots \cdot F(y_n)$$

where, $F(\cdot)$ is the cumulative distribution function, of each of the identically distributed random variables $Y_1, Y_2, \dots, Y_n$. In this model, there is no dependence between observations. In particular, for all $n \geq 1$ and all $y_1, y_2, \dots, y_n$.

$$P[Y_{n+h} \leq y | Y_1 \leq y_1, Y_2 \leq y_2, \dots, Y_n \leq y_n] = P[Y_{n+h} \leq y]$$

It shows that knowledge of $Y_1, Y_2, \dots, Y_n$ is of no value for predicting the behavior of Y_{n+h} . Given the values of $Y_1, Y_2, \dots, Y_n$ the function f that minimizes the mean squared error.

$E[Y_{n+h} - f(Y_1, Y_2, \dots, Y_n)]^2$ is in fact identically zero.

Although this means that IID noise is rather uninteresting process for forecasters, it plays an important role as a building block for more complicated time series model.

Example (Binary process):

Let us consider a sequence of iid random variables $\{Y_t; t = 1, 2, \dots\}$ with $P\{Y_t = 1\} = p$ and $P\{Y_t = -1\} = 1 - p$ where $p = \frac{1}{2}$.

The time series obtained by tossing a penny repeatedly and scoring +1 for each head and -1 for each tail is usually modeled as a realization of this process.

Random walk: The random walk $\{s_t, t=0, 1, 2, \dots, 5\}$ is obtained by cumulatively summing (or integrating) iid random variables. Thus a random walk with zero mean is obtained by defining $s_0 = 0$ and

$$s_t = \epsilon_1 + \epsilon_2 + \epsilon_3 + \dots + \epsilon_t; \quad t = 1, 2, \dots$$

where $\{\epsilon_t\}$ is IID noise. If $\{\epsilon_t\}$ is the binary process then $\{s_t; t=0, 1, 2, \dots, 5\}$ is called a simple symmetric random walk.

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

Definition: A random walk is defined as a process where the current value of a variable is composed of the past value plus an error term defined as a white noise (a normal variable with mean zero and variance one).

mathematically: $y_t = y_{t-1} + \epsilon_t$.

The best prediction of y for next period is the current value.

In other words, the process doesn't allow to predict the change ($y_t - y_{t-1}$)

That is, the change of y is absolutely random.

The mean of a random walk process is constant but its variance is not.

Therefore, a random walk process is nonstationary and its variance increases with t . In practice, the presence of a random walk process makes the forecast process very simple since all the future values of y_{t+s} for $s > 0$ is simply y_t .

Random Walk model: A random walk model in time series is a type of stochastic process where the value of a variable evolves over time in a way that each step or movement is random and unpredictable. Over time, the random errors accumulate, causing the time series to drift unpredictably. Sometimes a drift term is added, making the model:

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

where μ is a constant that represents the drift or the average

A drift behave like a trend.
The mathematical process has the following form:

$$y_t = \underbrace{y_0 + \alpha t}_{\text{deterministic trend}} + \underbrace{\left(\sum_{t=1}^n \epsilon_t \right)}_{\text{stochastic trend}}$$

If two time series are independent random walk processes then the relationship between the two doesn't have an economic meaning. If one still estimates a regression model between the two the following results are expected:

- ① High R^2
 - ② Low Durbin-Watson statistic (or high ρ)
 - ③ High t ratio for the slope coefficient.
- This indicates that the results from the regression are spurious.

Application:

- ① The random walk model is often used to model stock prices, exchange rates or other financial data where the future values are difficult to predict based on past values alone.

② It is a special case of the AR(1) model where the coefficient of the lagged variable is equal to 1.

A general approach to time series modeling:

- ① 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 behavior.
 - (d) Any outlying observation.

Remove the trend and seasonal ~~the~~ components to get stationary residuals (or irregular components). To achieve this goal, it may sometimes be necessary to apply a preliminary transformation of the data.

② For example, if the magnitude of the fluctuations appears to zero roughly linearly with the level of the series, then the transformed series $\{\ln Y_1, \ln Y_2, \dots, \ln Y_n\}$ will have fluctuations of more constant magnitude. There are several ways in

which trend and seasonality can be removed, some involving estimating the components and subtracting them from the data and others depending on differencing the data i.e. replacing the original series $\{Y_t\}$ by $\{u_t = Y_t - Y_{t-d}\}$ for some positive integer d . whichever method is used, the aim is to produce a stationary process.

③ choose a model to fit the residuals, making use of various sample statistics including the sample autocorrelation function
④ forecasting will be achieved by forecasting the residuals and then investing the transformations described above to arrive at forecasts of the original series $\{Y_t\}$.

Stationary process: A time series $\{Y_t; t=0, \pm 1, \pm 2, \dots\}$ is said to be stationary if it has similar statistical properties to the 'time shifted' series $\{Y_{t+h}; t=0, \pm 1, \pm 2, \dots\}$ for each integer h . Simply, a time series $\{Y_t; t=0, \pm 1, \pm 2, \dots\}$ is said to be stationary time series if it is independent of time t .

Non-stationary: A time series $\{Y_t; t=0, \pm 1, \pm 2, \dots\}$ is said to be non-stationary time series if it is ~~in~~ dependent of time t .

Weakly stationary: A time series $\{Y_t; t=0, \pm 1, \pm 2, \dots\}$ is said to be weakly stationary if its mean function and covariance function is independent of time t , i.e. -
(i) $\mu_Y(t)$ is independent of time t .
(ii) $\gamma_Y(t+h, t)$ is independent of time t for each h .

Usually weakly stationary is termed as simply stationary.

Strictly stationary: A time series $\{Y_t; t=0, \pm 1, \pm 2, \dots\}$ is said to be strictly stationary if the joint distribution of $Y_1, Y_2, \dots, Y_n$ is the same as the joint distribution of $Y_{1+h}, Y_{2+h}, \dots, Y_{n+h}$ for all integers h and $n > 0$, i.e.

$$\{Y_1, Y_2, \dots, Y_n\}^T \stackrel{d}{=} \{Y_{1+h}, Y_{2+h}, \dots, Y_{n+h}\}^T$$

strong stationary doesn't assume any distribution. It only says the probability

distribution should be the same. The distribution and joint distribution of a random variable is invariant to time.

(i) The strict stationarity property restricts 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.

Properties of Strictly Stationary time series.

$\{Y_t\}$:

The main properties of $\{Y_t\}$ are given below -

① The random variables Y_t are identically distributed.

② $\{Y_t, Y_{t+h}\}^T \stackrel{d}{=} \{Y_1, Y_{1+h}\}^T$; for all integers t and h .

(3) $\{Y_t\}$ is weakly stationary if $E\{Y_t^2\} < \infty$ for all t .

④ Weak stationarity doesn't imply strict stationarity.

⑤ Strict stationarity implies weak stationarity.

Ex Show that a strictly stationary process with $E\langle Y_t \rangle < \infty$ is weakly stationary or properties of weakly stationary time series.

Proof:

If $E\langle Y_t \rangle < \infty$ then the properties of strictly stationary are —

(i) the random variables Y_t are identical distributed.

(ii) $\langle Y_t, Y_{t+h} \rangle^T \stackrel{d}{=} \langle Y_1, Y_{1+h} \rangle^T$ for all integer t and h .

We can write $E\langle Y_t \rangle$ is independent of t and $\text{COV}\langle Y_t, Y_{t+h} \rangle = \text{COV}\langle Y_1, Y_{1+h} \rangle$, which is independent of t . Thus, $\langle Y_t \rangle$ is weakly stationary.

Weakly stationary time series:

- ① The distribution is defined by its mean and variance,
- ② The mean and variance is invariant to time shift.

A higher β makes the model more sensitive to changes in the trend, while a lower β smooths out trend changes.

Seasonal Smoothing (γ): (Triple Exponential Smoothing). In Holt-Winters's seasonal ~~smooth~~ smoothing, the γ hyperparameter controls the level of smoothing applied to the seasonal component of the time series. A higher γ adapts quickly to changing seasonal patterns, while a lower γ applies more smoothing, making the model slower to adapt to new seasonality trends.

By adjusting these hyperparameters carefully, we can control the level of smoothness and responsiveness of time series model, improving its forecasting performance.

Exponential Smoothing: Exponential smoothing is a time series forecasting technique used to make predictions by assigning exponentially decreasing weights to past observations. The key idea is that more recent observations are more relevant

for predicting future values. ~~There~~

A broadly accurate forecasting method for short-term forecasts. It assigns larger weights to more recent observations while assigning exponentially decreasing weights as the observations get increasingly distant. ~~It~~ It is the most effective when the time series parameters vary slowly over time.

Single exponential smoothing: It is used with univariate data with no trend and no seasonal components (pattern).

Univariate data: It involves only one variable or feature. Focuses on analyzing or forecasting a single variable over time or across observations. Easier to visualize and analyze compared to multivariate data.

<u>Year</u>	<u>Income</u>
1970	10000 Tk.
1971	50000 Tk.
1972	100000 Tk.

Multiplicative Data: It involves two or more variables or features. It is more complex to analyze because it requires understanding the relationship between the variables, often using techniques like multiplicative regression.

Single Exponential smoothing needs a single parameter called α ($0 < \alpha < 1$), α controls the rate at which the influence of past observations decreases exponentially.

Single exponential smoothing is a simple forecasting technique used to predict future values of a time series by smoothing past data points. It is most effective when the data has no clear trend or seasonality. The method applies a weighted average where more recent observations receive more weight and the influence of older observations decays exponentially.

The formula for single exponential

smoothing is -

$$F_{t+1} = \alpha \cdot Y_t + (1 - \alpha) F_t$$

where, F_{t+1} is the forecast for the next period -

y_t is the actual value at time t .
 f_t is the forecast for the current period.
 α is the smoothing parameter ($0 < \alpha < 1$), which controls how much weight is given to the most recent observation.

Double Exponential Smoothing: ~~Double~~ It is also known as Holt's linear trend method, is an extension of simple exponential smoothing that can account for both level and trend in the data. It is used when the data has a linear trend but no seasonal pattern. Two smoothing factors (α and β).

α controls the decay of the influence of change in trend. This method supports trends that change in additive ways (smoothing with linear trend) and trends that change in multiplicative ways (smoothing with exponential trend).

It uses two equations to update both the level and the trend.

Level equation: updates the smoothed value based on current observations and the trend.

Trend equation: updates the estimate of the trend.

Let F_t = forecast at time t .

T_t = estimated trend at time t .

S_t = smoothed level at time t .

Y_t = actual value at time t .

1. Level (smoothed value):

$$S_t = \alpha Y_t + (1 - \alpha)(S_{t-1} + T_{t-1})$$

2. Trend (smoothed trend):

$$T_t = \beta(S_t - S_{t-1}) + (1 - \beta)T_{t-1}$$

forecast for future period $t+k$:

$$F_{t+k} = S_t + kT_t$$

α controls the level smoothing and β controls the smoothing of the trend. It is effective for time series that exhibit upward or downward trends,

offering a more accurate forecast than single exponential smoothing for such cases.

Q What is training data and test data?
⇒ In time series analysis, training data and test data refer to the two segments of a dataset that are used for building and evaluating forecasting models. The goal is to train the model on one portion of the data and then test its performance on unseen data to ensure it can generalize well to future values.

Training Data: This is the portion of the time series dataset that is used to train the model. The model learns patterns, trends and relationships from this data. In time series forecasting, this data helps the model understand historical values to make future predictions. The training dataset typically includes a large portion (e.g. 70-80%) of the original data.

Ex: If we are forecasting monthly sales for 5 years (60 months), the training data might include the first 48 months.

Test data: This is the portion of the time series dataset that is used to evaluate the model's performance. After training, the model is tested on this unseen data to check how well it can predict future values. The test data usually includes the remaining portion of the dataset (e.g. 20-30%) and should represent real future time points that the model didn't see during training.

Ex: Continuing the previous example, the test data might include the last 12 months.

Training data allows the model to learn from past observations. Test data helps determine how well the model can predict future or unseen data, providing a measure of model accuracy.

Diff. Difference between single exponential smoothing and double exponential smoothing.

→ The primary difference between single exponential smoothing and double exponential smoothing lies in their ability to handle trends in the data.

Single exponential smoothing

① It doesn't handle trend.

② There is only one smoothing parameter which is α ($0 < \alpha < 1$)

③ It is best for such data with no trend and seasonality.

④ Predicting sales with no growth or decline.

Double exponential smoothing

① It handles trend component.

② There are two smoothing parameters such as α and β ($0 < \alpha < 1$ and $0 < \beta < 1$).

③ It is best for such data with a linear trend.

④ Predicting sales that show steady growth over time.

Triple Exponential Smoothing: It is also known as Holt-Winters method is an advanced forecasting technique used to predict future values of a time series that exhibit both trend and seasonality. This method extends Double Exponential Smoothing by incorporating a seasonal component. This technique applies exponential smoothing three times —

① Level smoothing — the smoothed estimate of the level or baseline value of the time series.

② Trend smoothing → the estimated trend (upward or downward) of the time series.

③ Seasonal smoothing → the repeating seasonal pattern in the data.

1. Level equation:

$$S_t = \alpha(Y_t - T_{t-s}) + (1-\alpha)(S_{t-1} + T_{t-1})$$

2. Trend equation

$$T_t = \beta(S_t - S_{t-1}) + (1-\beta)T_{t-1}$$

3. Seasonal equation:

$$I_t = \gamma(Y_t - S_t) + (1 - \gamma)I_{t-s}$$

forecast equation:

$$F_{t+k} = S_t + kT_t + I_{t-s+k}$$

I_t = seasonal component at time t .

α, β, γ = smoothing parameters.

γ controls the influence of the seasonal components.

Key-features of Triple exponential smoothing

(1) I_t can account for both upward or downward trends and regular seasonal patterns, making it suitable for complex time series data.

(2) Holt-Winter's exponential smoothing has two categories depending on the nature of the seasonal component

a) Holt-Winter's Additive method (for seasonality that is additive)

b) Holt-Winter's Multiplicative method (for seasonality that is multiplicative)

Researchers can choose between additive and multiplicative models based on the nature of the seasonality in the data.

b) It provides accurate forecasts by combining the level, trend and seasonal effects making it highly effective for business and other fields where seasonal trends are prevalent. Such as -

- ① monthly retail sales
- ② seasonal demand for products

How to configure exponential smoothing?

① Need to specify all the model parameters explicitly. This can be challenging for both beginners and experts.

② Numerical optimization is commonly used to search for and find the smoothing factors (α , β and γ) for the model resulting in the most negligible error.

(3) An exponential smoothing method can obtain values for unknown parameters by estimating them from the observed data.

(4) The initial values and unknown parameters can be estimated by minimizing the sum of the squared errors (SSE).

Exponential smoothing: For any fixed $\alpha \in [0, 1]$, the one-sided moving averages, $\hat{m}_t, t = 1, \dots, n$ defined by the recursions, $\hat{m}_t = \alpha X_t + (1 - \alpha) \hat{m}_{t-1}, t = 2, \dots, n$ and, $\hat{m}_1 = X_1$

the value of α . Application of this equation is often referred to as exponential smoothing. Since the recursions imply that for $t \geq 2$, $\hat{m}_t = \sum_{j=0}^{t-2} \alpha(1-\alpha)^j X_{t-j} + (1-\alpha)^{t-1} X_1$, a weighted moving average of $X_t, X_{t-1}, \dots$ with weights decreasing exponentially (except for the last one).

Benefits of exponential smoothing:

The moving average method weights all past observations equally when they fall within the moving average window and it gives observations outside the window zero

Eight
⇒ Simplicity: Exponential smoothing methods are relatively simple to understand and implement, making them accessible to researchers without extensive statistical backgrounds.

Flexibility: There are various forms of exponential smoothing (e.g. single, double and triple) allowing researchers to choose a method that fits the characteristics of their data such as trend and seasonality.

Adaptive weighting: The method assigns exponentially decreasing weights to past observations, meaning more recent data has a greater influence on forecasts. This is beneficial in capturing changes in trends and patterns.

Fewer data requirements: It requires less historical data compared to some other forecasting methods, making it suitable for situations where data is limited.

Testing the Estimated Noise Sequence;

The residuals are obtained by differencing the data or by estimating and subtracting the trend and seasonal components. If there is no dependence between these residuals, we can regard them as observations of independent random variables and there is no further modeling to be done except to estimate their mean and variance.

Now, we examine following simple tests for checking the hypothesis that the residuals are independent and identically distributed random variables.

It can be tested by the following test method:

- ① The Sample Auto-correlation function,
- ② The Portmanteau Test
- ③ The Turning Point test
- ④ The Difference sign test
- ⑤ The Rank test
- ⑥ Checking for normality.

The Sample Auto-correlation function: Let

$U_1, U_2, \dots, U_n$ be an iid sequence with finite variance. Then for large n , the sample autocorrelation function of the sequence

are approximately iid with

$$\hat{\rho} \sim N(0, \frac{1}{n}).$$

Let $u_1, u_2, \dots, u_n$ be the realization of the iid sequence. Now, about 95% of the sample autocorrelations ($\hat{\rho}$) should fall between $(\frac{-1.96}{\sqrt{n}}, \frac{1.96}{\sqrt{n}})$. If we compute $\hat{\rho}$ up to lag-40 as

$$\begin{array}{ccccccc} \text{lag-1} & \text{lag-2} & \dots & \text{lag-40} \\ (u_t, u_{t-1}) & (u_t, u_{t-2}) & \dots & (u_t, u_{t-40}) \\ u_1, u_1 & u_2, u_1 & \dots & u_{41}, u_1 \\ u_2, u_2 & u_3, u_2 & \dots & u_{42}, u_2 \\ \vdots & \vdots & & \vdots \\ u_n, u_{n-1} & u_n, u_{n-2} & \dots & u_n, u_{n-40} \end{array}$$

Decision Rule: If more than 2 or 3 values of $\hat{\rho}$ fall outside the bounds or that one value falls far outside the bounds, then we accept the null hypothesis, i.e., the population sequence is not iid otherwise reject the null hypothesis.

The Portmanteau test: To test we consider the test statistic as -

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

We know that

$$\hat{\rho} \sim N(0, \frac{1}{n}) \text{ i.e. } \sqrt{n} \hat{\rho} \sim N(0, 1)$$

Q is approximately distributed as the sum of sequences of the independent $N(0, 1)$ random variables $\sqrt{n} \hat{\rho}(j)$; $j=1, 2, \dots, h$ i.e., as Chi-squared with h degrees of freedom. For a large value of Q suggests that the sample autocorrelation of the data are too large for an iid sequence. We therefore reject the iid hypothesis at level α if

$Q > \chi^2_{1-\alpha}(h)$ where $\chi^2_{1-\alpha}(h)$ is the $1-\alpha$ quantile of the Chi-squared distribution with h degrees of freedom,

Refinement of the test:

(a) Ljung and Box (1978) modified Q and is defined as -

$$Q_{L.B} = n(n+2) \sum_{j=1}^h \frac{\hat{\rho}^2(j)}{n-j}$$

whose distribution is better approximated by the Chi-squared distribution with

(b) Another Portmanteau test, formulated by McLeod and Li (1983), can be used for testing whether the data are an iid sequence of normally distributed random variables. They defined the test statistic as -

$$\tilde{Q} = n(n+2) \sum_{k=1}^h \frac{\hat{\rho}_{ww}^{(k)}}{n-k}$$

where, $\hat{\rho}_{ww}^{(k)}$ is the sample autocorrelation of the squared data. The hypothesis of iid normal data is then rejected at level α if the observed value of $\tilde{Q}$ is larger than the $(1-\alpha)$ quantile of the $\chi^2_{1-\alpha}(h)$ distribution.

The Turning Point Test: If $u_1, u_2, \dots, u_n$ is a sequence of observations, we say that there is a turning point at time i , $1 < i < n$, if $u_i > u_{i+1}$ or if $u_{i-1} > u_i$ and $u_i < u_{i+1}$, i.e.

for any 3 points, there may be two turning point. So the probability of a turning point at time i is $\frac{2}{3}$. If T is the total number of turning

points of an iid sequence of length n , then the expected value of T is -

$$\mu_T = E(T) = \frac{2}{3}(n-2) \quad \left[\text{There may be } (n-2) \text{ turning points} \right]$$

$$\sigma_T^2 = \text{var}(T) = \left(\frac{16n-29}{90} \right)$$

A large value of $(T - \mu_T)$ indicates that the series is fluctuating more rapidly than expected for an iid sequence. On the other hand the value of $(T - \mu_T)$ much smaller indicates a positive correlation between neighbouring observations. For an iid sequence with $x_i \in \mathbb{R}$, T is approximately -

$$T \sim N(\mu_T, \sigma_T^2)$$

The test statistic is -

$$Z = \frac{|T - \mu_T|}{\sigma_T} \sim N(0, 1)$$

This means we can carry out a test of the iid hypothesis, rejecting it at level α if $Z > \Phi_{\frac{1-\alpha}{2}}$ where $\Phi_{\frac{1-\alpha}{2}}$ is the $\frac{1-\alpha}{2}$ quantile of the standard normal distribution.

The Difference Sign Test: Let $\{u_1, u_2, \dots, u_n\}$ be a sequence of observations where $u_i > u_{i-1}$ i.e. $u_i - u_{i-1}$ is positive. Let s be the number of total count between the difference $(u_i - u_{i-1})$. Assume that there is no zero difference in the sequence. If it shows a zero difference, then it drops from the analysis. Therefore, n is reduced. For an iid sequence, we have -

$$\mu_s = E(s) = \frac{1}{2}(n-1)$$

$$\text{And } \sigma_s^2 = \text{Var}(s) = \frac{n+1}{12}$$

If n is large then s is approximately $N(\mu_s, \sigma_s^2)$

The test statistic is -

$$Z = \frac{|s - \mu_s|}{\sigma_s} \sim N(0, 1)$$

A large positive (or negative) value of $s - \mu_s$ indicates the presence of an increasing (or decreasing) trend in the data. We therefore reject the assumption of no trend in the data if $Z > \Phi_{1-\alpha/2}$ where $\Phi_{1-\alpha/2}$ is the

$\frac{1-\alpha}{2}$ quantile of the standard normal distribution.

Rank Test: The Rank test is particularly useful for detecting a linear trend in the data. Define P be the number of pairs (i, j) such that $u_j > u_i$ and $j > i$, $i = 1, 2, \dots, (n-1)$. There is a total of $\binom{n}{2} = \frac{n}{2}(n-1)$ pairs (i, j) with $j > i$. For an iid sequence $u_1, u_2, \dots, u_n$, each event $\{u_j > u_i\}$ has probability $\frac{1}{2}$ and the mean of P is —

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

$$\text{and } \tilde{\sigma}_P^2 = \text{var}(P) = \frac{n(n-1)(2n+5)}{72}$$

If n is large then P is approximately $N(\mu_P, \tilde{\sigma}_P^2)$.

The test statistic is —

$$Z = \frac{P - \mu_P}{\tilde{\sigma}_P} \sim N(0, 1)$$

A large positive (or negative) value of $P - \mu_P$ indicates the presence of an increasing (decreasing) trend in the data. The assumption that a sample of sequence $u_1, u_2, \dots, u_n$ is an iid sequence is therefore

rejected at level α if $z > \Phi_{1-\alpha}^*$ where $\Phi_{1-\alpha}^*$ is the $\frac{1-\alpha}{2}$ quantile of the standard normal distribution.

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

Checking for Normality: If the noise process is Gaussian, i.e. if all of its joint distributions are normal, then stronger conclusions can be drawn when a model is fitted to the data. The following test enables us to check whether it is reasonable to assume that the observations from an iid sequence are also Gaussian. Let $U(1) < U(2) < \dots < U(n)$ be an ordered statistics of a random sample $U_1, U_2, \dots, U_n$ from the distribution $N(\mu, \sigma^2)$. If $Y(1) < Y(2) < \dots < Y(n)$ are the

order statistics
size n , then

$$E(U(j)) = \mu + \sigma T_j$$

where $T_j = E(Y(j))$; $j = 1, 2, \dots, n$. Thus a graph of the points $(T_1, U(1)) \dots (T_n, U(n))$ (known as a ω plot) should be approximately linear. However, if the sample values U_i are not normally distributed, then the graph should be nonlinear. Consequently, the squared correlation of the points $(T_i, U(i))$; $i = 1, 2, \dots, n$ should be near 1 if the normal assumption is correct. The assumption of normality is therefore rejected if the squared correlation R^2 is sufficiently small. If we approximate T_i by $\Phi^{-1}\left(\frac{i-0.5}{n}\right)$, then R^2 reduces to

$$R^2 = \frac{\left[\sum_{i=1}^n (U(i) - \bar{U}) \Phi^{-1}\left(\frac{i-0.5}{n}\right) \right]^2}{\sum_{i=1}^n (U(i) - \bar{U})^2 \sum_{i=1}^n \left[\Phi^{-1}\left(\frac{i-0.5}{n}\right) \right]^2}$$

where, $\bar{U} = \frac{U_1 + U_2 + \dots + U_n}{n}$

Percentage points for the distⁿ of R^2 , assuming normality of the sample values are given by Shapiro and Francia (1972) for sample sizes $n \leq 100$. For $n = 200$, $p(R^2 < 0.987) = 0.05$

What is Gaussian distribution?

→ A Gaussian distribution is a continuous probability distribution characterized by a symmetric, bell-shaped curve. The term ~~normal~~ normal distribution is more commonly used in applied statistics, while Gaussian distribution refers to the theoretical concept, named after the mathematician Carl Friedrich Gauss, who studied its properties.

Difference between white noise and IID noise.

White Noise

① White noise

means uncorrelated random variables often with constant mean and variance.

(2) White noise doesn't imply that every time step's noise comes from same probability dist'n, so, it might not be identically distributed across all time steps.

IID Noise

① IID noise is a stricter form, requiring each time point to be independent and identically distribution.

(2) IID noise assumes that every time step's noise comes from the same probability dist'n.

3) white noise is uncorrelated, meaning there's no predictable relationship between values at different time steps, but they don't need to be independent.

(a) In ARIMA models, the residuals are often modelled as white noise, implying no serial correlation between them.

5) white noise is defined by its frequency domain characteristics (flat power spectrum)

(a) white noise requires both no correlation and independence, which is a stronger assumption.

(a) It is often used in simpler statistical tests where independence is a strict requirement. IID noise can be from any dist'n (not necessarily Gaussian or zero-mean) and doesn't make any assumption about the frequency domain.

 The frequency domain refers to the analysis of signals with respect to frequency, rather than time. In contrast to the time domain, where we analyze how a signal behaves over time, the frequency domain focuses

on how much of the signal lies within each given frequency range.

One of the main tools for converting a time domain signal into the frequency domain is the Fourier transform. It expresses the signal as a sum of sinusoids with different frequencies and amplitudes.

The frequency domain is widely used to analyze and filter signals. It is crucial for understanding system stability and designing filters.

Difference between random walk with and without drift.

⇒ A random walk is a statistical model where the current value of a variable depends on its previous value plus a random error term. When we talk about a random walk "with drift" versus "without drift", we are referring to whether or not there is a constant trend added to the movement of the

random walk over time.

Random walk without drift: In a random walk without drift, the next value of the variable is simply the previous value plus a random error term (noise).
mathematical form:

$$X_t = X_{t-1} + \epsilon_t$$

where, X_t is the current value, X_{t-1} is the previous value, and ϵ_t is the random error term, often assumed to be normally distributed with a mean zero. Over time, the value of the series can fluctuate in any direction (up or down), but it has no inherent trend. The expected change is zero, meaning the series doesn't drift in any consistent direction.

Ex: stock prices often follow a random walk without drift in the short term because their movements are largely unpredictable and fluctuate around a constant average.

Random walk with drift: A random walk with drift includes a constant trend or drift term. This means that, in

In addition to the random error term, there is a systematic movement in one direction (upward or downward) in the mathematical form:

$$X_t = X_{t-1} + \mu + \epsilon_t$$

where μ is the drift term (a constant). Over time, the series tends to trend in the direction of the drift μ . If $\mu > 0$ upward, if $\mu < 0$, downward. The random fluctuations still occur, but the series as a whole exhibits a systematic trend due to the drift term.

Example: A company's stock price might exhibit a random walk with drift, if over time the company grows consistently, leading to an upward trend in its stock price despite day-to-day fluctuations.

Random walk without drift

(1) There is no drift term; the next value depends on the previous value and random noise.

Random walk with drift

(2) There is a drift term μ , which adds a constant trend to the random walk.

(2) The expected change in the variable is zero

(3) The expected change is $\neq$ (either positive or negative) leading to a consistent upward or downward trend over time

 First-order moving MAC(1) process:

Consider the series defined by the ~~algorithm~~ equation -

$$X_t = Z_t + \theta Z_{t-1}; t = 0, \pm 1, \dots$$

where $Z_t \sim WN(0, \sigma^2)$ and θ is a real-valued constant

$$E(X_t) = 0, E X_t^2 = \sigma^2(1 + \theta^2) < \infty \text{ and}$$

$$\gamma_X(t+h, t) = \begin{cases} \sigma^2(1 + \theta^2) & \text{if } h = 0 \\ \sigma^2\theta & \text{if } h = \pm 1 \\ 0 & \text{if } |h| > 1 \end{cases}$$

$\{X_t\}$ is stationary. The autocorrelation function of $\{X_t\}$ is -

$$\rho_X(h) = \begin{cases} 1 & \text{if } h = 0 \\ \theta / (1 + \theta^2) & \text{if } h = \pm 1 \\ 0 & \text{if } |h| > 1 \end{cases}$$

First-order Autoregressive process AR(1) process:

Let us assume now that $\{X_t\}$ is stationary series satisfying the equations —

$$X_t = \phi X_{t-1} + Z_t \quad \text{--- (1)}$$

where $\{Z_t\} \sim WN(0, \sigma^2)$, $|\phi| < 1$, and Z_t is uncorrelated with X_t for each $s < t$.

Using the fact that $E Z_t = 0$, ^{we} see at once that $E X_t = 0$.

To find the autocorrelation function of $\{X_t\}$, we multiply each side of (1) by X_{t-n} ($n > 0$) and then take expectations to get —

$$\begin{aligned} \gamma_X(n) &= \text{Cov}(X_t, X_{t-n}) \\ &= \text{Cov}(\phi X_{t-1} + Z_t, X_{t-n}) + \text{Cov}(Z_t, X_{t-n}) \\ &= \phi \gamma_X(n-1) + 0 \\ &= \dots \phi^n \gamma_X(0) \end{aligned}$$

Observing that $\gamma(n) = \gamma(-n)$.

$$\rho_X(n) = \frac{\gamma_X(n)}{\gamma_X(0)} = \phi^{|n|} ; n = 0, \pm 1, \dots$$

It follows from the linearity of the covariance function in each of its arguments and the fact that z_t is uncorrelated with x_{t-1} that

$$\begin{aligned}\gamma_x(0) &= \text{Cov}(x_t, x_t) \\ &= \text{Cov}(\phi x_{t-1} + z_t, \phi x_{t-1} + z_t) \\ &= \phi^2 \gamma_x(0) + \sigma^2\end{aligned}$$

and hence that $\gamma_x(0) = \frac{\sigma^2}{1-\phi^2}$

Let $x_1 \dots x_n$ 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 -

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

The sample autocorrelation function is -

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

Methods of calculating ACF:

① Direct method

② Sample autocovariance function

③ ACF plot

④ Yule Walker equations.

(5) fast fourier transform.

(6) Parzen's modeling.

(7) cross-covariance function.

Q. What is linear process? How to transform a second order stationary process to a linear process?

→ The time series $\{X_t\}$ is a linear process if it has the representation

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j} \text{ for all } t.$$

where, $\{Z_t\} \sim WN(0, \sigma^2)$ and $\{\psi_j\}$ is a sequence of constants with

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty.$$

Every second order stationary process is either a linear process or can be transformed to a linear process by subtracting a deterministic component (Wold's decomposition).

Q. Why second order property is important?

(1) It helps to check for stationarity in the data.

(2) It plays a vital role in selecting

appropriate models in time series. Having an accurate model allows for better forecasting performance.

(3) Used to estimate model parameters efficiently.

(4) Used to identify heteroscedasticity which refers to the varying levels of variability in a time series.

(5) Essential for model diagnostic checking.

(6) Helps to identify outlier or unusual observation in the data.

Autoregressive models:

features -

(1) No consistent trend (upward or downward) over the entire time span.

(2) ~~no~~ No seasonality

(3) ~~no~~ No outliers

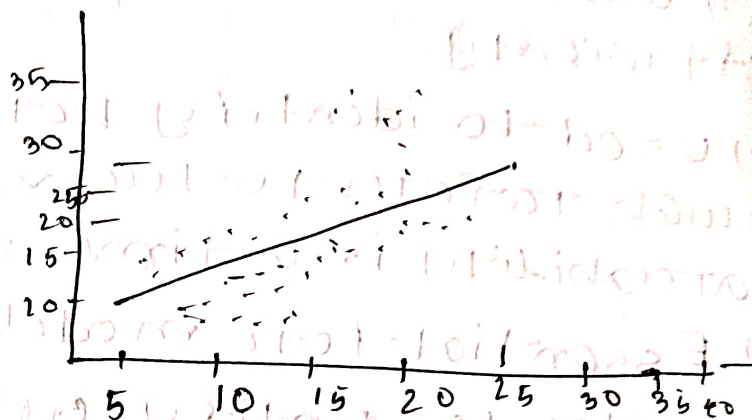
(4) Difficult to judge whether the variance is constant or not

Then AR(1) model might be appropriate.

Justification:

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

t	x_t	x_{t-1}
1	13	—
2	14	13
3	8	14
4	10	8
5	16	10



→ a moderately strong relationship

AR models: An autoregressive

model is when a value from a time series is regressed on previous values from that same time series.

For example,

y_t on y_{t-1}

$$y_t = \rho_0 + \rho_1 y_{t-1} + \epsilon_t$$

$$x_t = \phi x_{t-1} + z_t$$

AR(1) or 1st order autoregressive model.

$$y_t = \rho_0 + \rho_1 y_{t-1} + \rho_2 y_{t-2} + \epsilon_t$$

$$\text{or } x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + z_t$$

AR(2)

AR(P) model will be

$$x_t = \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + z_t$$

MA(n) model :

$$x_t = \theta_1 z_{t-1} + \theta_2 z_{t-2} + \dots + \theta_n z_{t-n} + z_t$$

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

Autoregressive moving Average (ARMA) models -
modeling financial time series.

The merger between AR(P) and MA(n)
models -

AR(P)

+

MA(n)

Explains the
momentum and mean
reversion effects
often observed in
trading markets
market participant
effect)

+

Captures the shock
effects (unexpected
events affecting
the observation
process) 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 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.

$$\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$$

$$\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$$

Autoregressive integrated moving Average (ARIMA) models -

- ① An upward trend (possibly long-term; spikes) possibly a curved one.
- ② Seasonality - a regularly repeating pattern of highs and lows related to quarters of the year.
- ③ No obvious outliers.

(4) Increasing variation as we move across time, although that's uncertain.

ARIMA methods for dealing with series that exhibit both trend and seasonality could be used here.

③ Another option, using ordinary regression methods.

A stationary solution of the ARMA(1,1) equation exists if and only if $\phi \neq \pm 1$.

If $|\phi| < 1$, then the unique stationary solution is

$$X_t = Z_t + (\phi + \theta) \sum_{j=1}^{\infty} \phi^{j-1} Z_{t-j}.$$

In this case, we say that $\{X_t\}$ is causal or a causal function of $\{Z_t\}$. Since

X_t can be expressed in terms of the current and past values of Z_s , $s \leq t$.

If $|\phi| > 1$, then the unique solution is -

$$X_t = \chi(\theta)\theta(\phi)Z_t = -\theta\phi^{-1}Z_t - (\theta + \phi) \sum_{j=1}^{\infty} \phi^{-j-1} Z_{t+j}$$

The solution is non-causal, since X_t is then a function of Z_s , $s \geq t$.

3rd Tutorial

Spectral Analysis

Time Domain Analysis: The variable value of a variable Y_t at time t has typically been described in terms of a sequence of innovations

$\{ \epsilon_t \}$ in model of the form -

$$Y_t = \mu + \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$$

The focus has been on the implication of such a representation of the covariance between Y_t and Y_s at distinct dates t and s . This is known as analyzing the properties of $\{Y_t\}$ in the time domain.

Spectral Analysis: Instead the value of Y_t can be described as a weighted sum of period functions of the form $\cos(\omega t)$ and $\sin(\omega t)$, where ω denote as particular frequency.

$$Y_t = \mu + \int_0^{\pi} \cos(\omega t) \alpha(\omega) d\omega + \int_0^{\pi} \beta(\omega) \sin(\omega t) d\omega$$

The goal will be to determine how important cycles of different frequencies are in accounting for the behaviour of Y . It is known as frequency domain or spectral analysis.

The summability of $|\gamma(\cdot)|$ implies that the series in (1) converges absolutely. Since,

$$|e^{ihn}|^2 = \cos^2(h\lambda) + \sin^2(h\lambda) = 1$$

Since, $\cos$ and $\sin$ have period 2π . So, also f converges and it suffices to confine attention to the values of f on the interval $(-\pi, \pi]$.

Basic properties of Spectral Density:

- 1) f is even, i. e. $f(\lambda) = f(-\lambda)$ [Symmetry]
- 2) $f(\lambda) \geq 0$ for all $\lambda \in (-\pi, \pi)$ [Non-negativity]
- 3) $\gamma(k) = \int_{-\pi}^{\pi} e^{ik\lambda} f(\lambda) d\lambda = \int_{-\pi}^{\pi} \cos(k\lambda) f(\lambda) d\lambda$.
[Existence of ACF]

Proof 1:

Let $\lambda \in \mathbb{R}$,

$$\begin{aligned} f(\lambda) &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ihn\lambda} \gamma(h) \\ &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} [\cos(h\lambda) - i\sin(h\lambda)] \gamma(h) \\ &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \cos(h\lambda) \gamma(h) \quad [\text{Since } \sin(w) = -\sin(-w)] \\ &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \cos(-h\lambda) \gamma(h) \\ &= f(-\lambda). \end{aligned}$$

Proof 2: For each positive integer N define.

$$\begin{aligned}
 f_N(\lambda) &= \frac{1}{2\pi N} E \left[\left| \sum_{\pi=1}^N X_{\pi} e^{-i\pi\lambda} \right|^2 \right] \\
 &= \frac{1}{2\pi N} E \left[\sum_{\pi=1}^N X_{\pi} e^{-i\pi\lambda} \sum_{s=1}^N X_s e^{-is\lambda} \right] \\
 &= \frac{1}{2\pi N} \sum_{|h| < N} (N - |h|) e^{-ih\lambda} \gamma(h)
 \end{aligned}$$

where,

$$T_N = [\gamma(i-j)]_{i,j=1}^N$$

Clearly the function f_N is non-negative for each N and since

$$f_N(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ih\lambda} \gamma(h)$$

$$= f(\lambda) \text{ as } N \rightarrow \infty$$

Also, f must be non-negative i. e.

$$f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ih\lambda} \gamma(h) \geq 0$$

$$= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \cos(h\lambda) \gamma(h) \geq 0$$

Since, $\cos$ and $\gamma(h)$ is even function.

Proof 3:

We have,

$$\begin{aligned} \int_{-\pi}^{\pi} e^{ik\lambda} f(\lambda) d\lambda &= \int_{-\pi}^{\pi} \frac{1}{2\pi} e^{ik\lambda} \sum_{n=-\infty}^{\infty} e^{-in\lambda} \delta(\lambda) d\lambda \\ &= \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} \int_{-\pi}^{\pi} e^{i(k-n)\lambda} \delta(\lambda) d\lambda \\ &= \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} \delta(n) \int_{-\pi}^{\pi} \{ \cos[\lambda(k-n)] - i \sin[\lambda(k-n)] \} d\lambda \end{aligned}$$

for $k = n$.

$$\begin{aligned} &\int_{-\pi}^{\pi} \{ \cos[\lambda(k-n)] - i \sin[\lambda(k-n)] \} d\lambda \\ &= \int_{-\pi}^{\pi} (\cos 0 - i \sin 0) d\lambda = 2\pi \end{aligned}$$

for $k \neq n$

$$\begin{aligned} &\int_{-\pi}^{\pi} \{ \cos[\lambda(k-n)] - i \sin[\lambda(k-n)] \} d\lambda \\ &= \left[\frac{\sin[\lambda(k-n)]}{k-n} - i \frac{\cos[\lambda(k-n)]}{k-n} \right]_{\lambda=-\pi}^{\lambda=\pi} \end{aligned}$$

$$\begin{aligned} &= \frac{1}{k-n} \{ \sin[\pi(k-n)] + \sin[-\pi(k-n)] - \\ &\quad i \cos[\pi(k-n)] + i \cos[-\pi(k-n)] \} \end{aligned}$$

Since the difference between the frequencies $\pi(k-n)$ and $-\pi(k-n)$ is $2\pi(k-n)$ which is an integer multiple of 2π . Since the sine

and cosine-functions are periodic the magnitude of this integral is zero. Hence, only the term for $k = n$ in the sum in (*) is non-zero. Thus,

$$\int_{-\pi}^{\pi} e^{ikx} f(x) dx = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} \int_{-\pi}^{\pi} e^{i(k-n)x} \gamma(x) dx$$

$$= \gamma(k)$$

We also show that

$$\gamma(k) = \int_{-\pi}^{\pi} \cos(kx) f(x) dx.$$

Since, $f(x)$ is symmetric around $x = 0$,

$$\begin{aligned} \gamma(k) &= \int_{-\pi}^{\pi} e^{ikx} f(x) dx \\ &= \int_{-\pi}^0 e^{ikx} f(x) dx + \int_0^{\pi} e^{ikx} f(x) dx \\ &= \int_0^{\pi} e^{-ikx} f(-x) dx + \int_0^{\pi} e^{ikx} f(x) dx \\ &= \int_0^{\pi} e^{-ikx} f(x) dx + \int_0^{\pi} e^{ikx} f(x) dx \\ &= \int_0^{\pi} (e^{-ikx} + e^{ikx}) f(x) dx \\ &= 2 \int_0^{\pi} \cos(kx) f(x) dx \\ &= \int_0^{\pi} \cos(kx) f(x) dx + \int_0^{\pi} \cos(kx) f(x) dx \\ &= \int_{-\pi}^0 \cos(-kx) f(-x) dx + \int_0^{\pi} \cos(kx) f(x) dx \end{aligned}$$

$$= \int_{-\pi}^0 \cos(k\lambda) f(\lambda) d\lambda + \int_0^{\pi} \cos(k\lambda) f(\lambda) d\lambda$$

$$= \int_{-\pi}^{\pi} \cos(k\lambda) f(\lambda) d\lambda$$

Find the spectral density from the following processes

- (a) white noise
- (b) AR(1) process
- (c) MA(1) process
- (d) ARMA(1,1) process

For white noise:

Let $\{X_t\} \sim WN(0, \sigma^2)$, then

$\gamma(0) = \sigma^2$ and $\gamma(h) = 0$ for all $|h| > 0$. We know

that the spectral density of the auto-covariance function $\gamma(h)$ of a stationary time series is -

$$f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) e^{-i\lambda h} \quad \lambda \in (-\pi, \pi]$$

for white noise process.

$$f(\lambda) = \frac{1}{2\pi} \sigma^2 e^0$$

$$= \frac{\sigma^2}{2\pi} \quad \text{since } \gamma(h) = 0 \text{ for } |h| > 0$$

$$-\pi \leq \lambda \leq \pi$$

A process with this density is called white noise since each frequency in the

spectrum contributes equally to the variance of the process.

For AR(1) process:

we have,

$$X_t = \phi X_{t-1} + Z_t$$

where, $Z_t \sim WN(0, \sigma^2)$,

we know that, $\gamma(0) = \frac{\sigma^2}{1-\phi^2}$

$\gamma(h) = \phi^h \gamma(0)$. Thus, the spectral density of $\{X_t\}$ is given by -

$$f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ih\lambda} \gamma(h) ; -\infty \leq \lambda \leq \infty$$

Here, we get -

$$\begin{aligned} f(\lambda) &= \frac{1}{2\pi} \left[\gamma(0) + \sum_{h=1}^{\infty} (e^{-ih\lambda} + e^{ih\lambda}) \phi^h \gamma(0) \right] \\ &= \frac{\gamma(0)}{2\pi} \left[1 + \sum_{h=1}^{\infty} (\phi e^{-i\lambda})^h + \sum_{h=1}^{\infty} (\phi e^{i\lambda})^h \right] \\ &= \frac{\gamma(0)}{2\pi} \left[1 + \frac{\phi e^{-i\lambda}}{1 - \phi e^{-i\lambda}} + \frac{\phi e^{i\lambda}}{1 - \phi e^{i\lambda}} \right] \\ &= \frac{\sigma^2}{2\pi(1-\phi^2)} \left[\frac{(1 - \phi e^{-i\lambda})(1 - \phi e^{i\lambda}) + \phi e^{-i\lambda}(1 - \phi e^{i\lambda}) + \phi e^{i\lambda}(1 - \phi e^{-i\lambda})}{(1 - \phi e^{-i\lambda})(1 - \phi e^{i\lambda})} \right] \\ &= \frac{\sigma^2}{2\pi(1-\phi^2)} \left[\frac{1 - \phi^2}{1 - \phi(e^{-i\lambda} + e^{i\lambda}) + \phi^2} \right] \end{aligned}$$

$$2\cos k = e^{-ik} + e^{ik}$$

$$= \frac{\sigma^2}{2\pi} (1 - 2\theta \cos k + \theta^2)^{-1}$$

For π MA(1) process:

we have,

$$X_t = Z_t + \theta Z_{t-1}; t = 0, \pm 1, \pm 2, \dots$$

where, $Z_t \sim WN(0, \sigma^2)$ and θ is a real value constant. we know that -

$$\gamma(0) = \sigma^2(1 + \theta^2)$$

$$\text{and } \gamma(h) = \begin{cases} \theta \sigma^2 & ; h = \pm 1 \\ 0 & ; |h| > 1 \end{cases}$$

we know that the spectral density of $\{X_t\}$ is given by -

$$f(k) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ikh} \gamma(h); -\infty \leq k \leq \infty$$

$$= \frac{1}{2\pi} \left[\gamma(0) + \sum_{h=1}^{\infty} e^{-ikh} \gamma(h) + 0 \right]$$

$$= \frac{1}{2\pi} \left[\sigma^2(1 + \theta^2) + \theta \sigma^2 (e^{-ik} + e^{ik}) \right]$$

$$= \frac{\sigma^2}{2\pi} \left[1 + \theta^2 + \theta (e^{-ik} + e^{ik}) \right]$$

$$= \frac{\sigma^2}{2\pi} [1 + 2\theta \cos k + \theta^2]$$

$$\boxed{\begin{aligned} & - \theta e^{-ik} - \theta e^{ik} + \theta^2 \\ & 1 - \theta (e^{-ik} + e^{ik}) + \theta^2 \end{aligned}}$$

ARMA(1,1):

$$x_t - \phi x_{t-1} = z_t + \theta z_{t-1}$$

$$\gamma(n) = \begin{cases} \sigma^2 \left[1 + \frac{(\phi + \theta)^2}{1 - \phi^2} \right]; & n = 0 \\ \sigma^2 \left[(\phi + \theta) \phi^{n-1} + \frac{(\phi + \theta)^2 \phi^n}{1 - \phi^2} \right]; & |n| > 0 \end{cases}$$

$$\begin{aligned} f(\lambda) &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ihn} \gamma(n) \\ &= \frac{1}{2\pi} \left[\gamma(0) + \sum_{h=1}^{\infty} (e^{-ihn} + e^{ihn}) \gamma(n) \right] \\ &= \frac{\sigma^2}{2\pi} \left[\frac{1 + \theta^2 + 2\theta \cos \lambda}{1 - \phi^2 - 2\phi \cos \lambda} \right] \end{aligned}$$

ARMA(p,q)

$$\phi(p)x_t = \theta(q)z_t$$

$$f(\lambda) = \frac{\sigma^2}{2\pi} \cdot \frac{|\theta e^{-i\lambda}|^2}{|\phi e^{-i\lambda}|^2}$$

Periodogram Analysis: Periodogram analysis is a

method of determining the period of the cyclical term C_t in a time series of the form

$$Y_t = C_t + U_t$$

where U_t is a random fluctuation about C_t .

The cyclical term is represented as a sum of sine and cosine terms so that Y_t becomes

$$Y_t = \sum [A_i \cos(\omega_i t) + B_i \sin(\omega_i t)] + U_t.$$

For many series, there may be several periodic terms present with unknown periods and so not only the coefficients $\{A_i\}$, $\{B_i\}$ have to be estimated but also the unknown frequencies $\{\omega_i\}$. The so-called hidden periodicities can often be detected by the periodogram which is a plot of $I(\omega)$ against ω .

$$I(\omega) = \frac{2}{N} \left[\left\{ \sum_{t=1}^N Y_t \cos(\omega t) \right\}^2 + \left\{ \sum_{t=1}^N Y_t \sin(\omega t) \right\}^2 \right]$$

$$\text{where, } \omega = \frac{2\pi p}{N}; \quad p = 1, 2, \dots, \left[\frac{N}{2} \right]$$

N is the length of time series.

where, $[x]$ denotes the largest integer less than or equal to x . Large ordinates on this

Plot indicate the presence of a cyclical component at a particular frequency.

Example:

The ARMA(1,1) process,

$$X_t = Z_t + (\phi + \theta) \sum_{j=1}^{\infty} \phi^{j-1} Z_{t-j}$$

The autocovariance function for a causal time series is -

$$\gamma(h) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|h|}$$

$$\text{where } \psi_j = (\phi + \theta) \phi^{j-1}$$

Case $h = 0$

$$\text{cov}(X_t, X_t) = E[X_t, X_t] = \gamma(0) = \sigma^2 \sum_{j=0}^{\infty} \psi_j^2$$

$$E[X_t, X_t] = E\left[\left(Z_t + (\phi + \theta) \sum_{j=1}^{\infty} \phi^{j-1} Z_{t-j}\right)^2\right]$$

$$= E[Z_t^2] + E\left[(\phi + \theta) \sum_{j=1}^{\infty} \phi^{j-1} Z_{t-j}\right]^2 +$$

$$2(\phi + \theta) \sum_{j=1}^{\infty} \phi^{j-1} E(Z_{t-j}, Z_t)$$

Since, Z_t is a white noise,

$$E(Z_{t-j}, Z_t) = 0$$

$$\therefore E[X_t, X_t] = E[Z_t^2] + (\phi + \theta)^2 \sum_{j=1}^{\infty} \phi^{2j-2} E[Z_{t-j}^2]$$

$$= \sigma^2 + (\phi + \theta)^2 \sum_{j=1}^{\infty} \phi^{2j-2} \sigma^2$$

$$\gamma(0)$$

$$1 + x + x^2 + \dots = \frac{1}{1-x}$$

$$= \sigma^2 \left[1 + (\phi + \theta)^2 \sum_{j=1}^{\infty} \phi^{2j-2} \right]$$

$$= \sigma^2 \left[1 + (\phi + \theta)^2 (1 + \phi^2 + \phi^4 + \dots) \right]$$

$$= \sigma^2 \left[1 + \frac{(\phi + \theta)^2}{1 - \phi^2} \right] \quad \left[\because 1 + \phi^2 + \phi^4 + \dots = (1 - \phi^2)^{-1} \right]$$

where,

$$\sum_{j=1}^{\infty} \phi^{2j-2} = \sum_{j=0}^{\infty} \phi^{2j} = \sum_{j=0}^{\infty} (\phi^2)^j \text{ is geometric}$$

series converging to $\frac{1}{1 - \phi^2}$

Case $h=1$:

$$\text{cov}(X_t, X_{t-1}) = E[X_t, X_{t-1}] = \gamma(1)$$

$$= \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+1}$$

$$E[X_t X_{t-1}] = E \left[\left\{ z_t + (\phi + \theta) \sum_{j=1}^{\infty} \phi^{j-1} z_{t-j} \right\} \right.$$

$$\left. \left\{ z_{t-1} + (\phi + \theta) \sum_{j=2}^{\infty} \phi^{j-2} z_{t-j} \right\} \right]$$

$$= E \left[\left\{ z_t + z_{t-1}(\phi + \theta) + (\phi + \theta) \sum_{j=2}^{\infty} \phi^{j-1} z_{t-j} \right\} \right.$$

$$\left. \left\{ z_{t-1} + (\phi + \theta) \sum_{j=2}^{\infty} \phi^{j-2} z_{t-j} \right\} \right]$$

$$= E \left[\left\{ z_t + z_{t-1}(\phi + \theta) + (\phi + \theta)\phi \sum_{j=2}^{\infty} \phi^{j-2} z_{t-j} \right\} \right.$$

$$\left. \left\{ z_{t-1} + (\phi + \theta) \sum_{j=2}^{\infty} \phi^{j-2} z_{t-j} \right\} \right]$$

$$= (\phi + \theta) E[z_{t-1}] + (\phi + \theta)\phi \sum_{j=2}^{\infty} \phi^{j-2} E[z_{t-j}]$$

$$= (\phi + \theta)\sigma^2 + (\phi + \theta)\phi \sum_{j=2}^{\infty} \phi^{j-2} \sigma^2$$

$$= \sigma^2 (\phi + \theta) + (\phi + \theta)\phi \sum_{j=2}^{\infty} \phi^{j-2}$$

Case $h=2$:

$$\text{Cov}(X_t, X_{t-2}) = E[X_t X_{t-2}] = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j-2}$$

$$E[X_t X_{t-2}] = E\left[\left(z_t + (\phi + \theta) \sum_{j=1}^{\infty} \phi^{j-1} z_{t-j}\right) \left(z_{t+2} + (\phi + \theta) \sum_{j=2}^{\infty} \phi^{j-2} z_{t-j}\right)\right]$$

$$= E\left[z_t + z_{t+1}(\phi + \theta) + z_{t+2} \phi(\phi + \theta) + (\phi + \theta) \phi \sum_{j=3}^{\infty} \phi^{j-3} z_{t-j}\right] \left(z_{t+2} + (\phi + \theta) \sum_{j=2}^{\infty} \phi^{j-2} z_{t-j}\right)$$

$$= (\phi + \theta) \phi E[z_{t-2}] + (\phi + \theta)^2 \phi \sum_{j=3}^{\infty} \phi^{j-3} E[z_{t-j}]$$

$$= \sigma^2 ((\phi + \theta) \phi) + \frac{(\phi + \theta)^2 \phi^2}{1 - \phi^2} \sigma^2$$

$$= \sigma^2 \left[\phi(\phi + \theta) + \frac{(\phi + \theta)^2 \phi^2}{1 - \phi^2} \right]$$

we can see the pattern which as follows—

$$\gamma(h) = \sigma^2 \left[1 + \frac{(\phi + \theta)^2}{1 - \phi^2} \right] \text{ if } h=0$$

$$\gamma(h) = \sigma^2 \left[(\phi + \theta) \phi^{h-1} + \frac{(\phi + \theta)^2 \phi^h}{1 - \phi^2} \right] \text{ if } h > 0$$

Estimation of μ :

The moment estimator of the mean μ of a stationary process is the sample mean.

$$\bar{x}_n = \frac{x_1 + x_2 + \dots + x_n}{n}$$

It is an unbiased estimator of μ . since

$$E(\bar{x}_n) = \frac{E(x_1) + E(x_2) + \dots + E(x_n)}{n} = \mu$$

The mean squared error of $\bar{x}_n$ is -

$$E[(\bar{x}_n - \mu)^2] = \text{var}(\bar{x}_n)$$

$$= \frac{\text{var}(x_1 + x_2 + \dots + x_n)}{n^2}$$

$$= \frac{\text{var}[x_1] + \text{var}[x_2] + \dots + \text{var}[x_n] + \sum_{i \neq j} \text{cov}(x_i, x_j)}{n^2}$$

$$= \frac{\sum_{i=1}^n \sum_{j=1}^n \text{cov}(x_i, x_j)}{n^2}$$

$$\text{cov}(x_i, x_j)$$

$$\text{lag } h = i - j$$

$x(0)$ repeat n times

$x(-1)$ " $(n-1)$ times

$x(1)$ " $(n-1)$ "

$x(2)$ " $(n-2)$ "

$x(-2)$ " $(n-2)$ times

$$\text{var}[\bar{x}_n] = \frac{n\sigma(0) + (n-1)\sigma(-1) + (n-1)\sigma(1) + (n-2)\sigma(-2) + \dots + \sigma(-(n-1)) + \sigma(n-1)}{n^2}$$

$$= \frac{\sigma(0) + \left(\frac{n-1}{n}\right)\sigma(-1) + \left(\frac{n-1}{n}\right)\sigma(1) + \left(\frac{n-2}{n}\right)\sigma(-2) + \left(\frac{n-2}{n}\right)\sigma(2) + \dots + \frac{1}{n}\sigma(-(n-1)) + \frac{1}{n}\sigma(n-1)}{n}$$

Note that,

$$\frac{n-1}{n} = 1 - \frac{1}{n}$$

$$\frac{n-2}{n} = 1 - \frac{2}{n}$$

$$\frac{1}{n} = 1 - \frac{n-1}{n}$$

$$= n^{-1} \left[\sigma(0) + \left(1 - \frac{1}{n}\right)\sigma(-1) + \left(1 - \frac{1}{n}\right)\sigma(1) + \left(1 - \frac{2}{n}\right)\sigma(-2) + \left(1 - \frac{2}{n}\right)\sigma(2) + \dots + \left(1 - \frac{n-1}{n}\right)\sigma(n-1) \right]$$

$$= n^{-1} \sum_{h=-(n-1)}^{n-1} \left(1 - \frac{|h|}{n}\right) \sigma(h)$$

$$= \frac{\sum_{h=-n}^n \left(1 - \frac{|h|}{n}\right) \sigma(h)}{n}$$

$$\gamma(h) = \begin{cases} 1 & \text{if } h = 0 \\ -0.5 & \text{if } h = \pm 2 \\ -0.25 & \text{if } h = \pm 3 \\ 0 & \text{; otherwise} \end{cases}$$

$$f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ih\lambda} |\gamma(h)|$$

$$= \frac{1}{2\pi} [1 - 0.5(e^{-2i\lambda} + e^{2i\lambda}) - 0.25(e^{-3i\lambda} + e^{3i\lambda})]$$

$$= \frac{1}{2\pi} [1 - \cos 2\lambda - 0.5 \cos 3\lambda] \leq 0 \text{ at } \lambda = 0$$

$\gamma(h)$ can't be an autocovariance function.

Spectral density of an MA(1) process

$$X_t = Z_t + \theta Z_{t-1}$$

Since, $\{Z_t\} \sim WN(0, \sigma^2)$. Then

$$f(\lambda) = \frac{\sigma^2}{2\pi} [1 + \theta^2 + \theta(e^{-i\lambda} + e^{i\lambda})]$$

$$= \frac{\sigma^2}{2\pi} [1 + 2\theta \cos \lambda + \theta^2]$$

Spectral density of an AR(1) process

$$X_t = \phi X_{t-1} + Z_t$$

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

$$f(\lambda) = \frac{\sigma^2}{2\pi(1-\phi^2)} \left[1 + \sum_{h=1}^{\infty} \phi^h (e^{-ih\lambda} + e^{ih\lambda}) \right]$$

$$= \frac{\sigma^2}{2\pi(1-\phi^2)} \left[1 + \frac{\phi e^{i\lambda}}{1-\phi e^{i\lambda}} + \frac{\phi e^{-i\lambda}}{1-\phi e^{-i\lambda}} \right]$$

$$= \frac{\sigma^2}{2\pi} [1 - 2\phi \cos \lambda + \phi^2]^{-1} \quad \sigma^2 = 1$$

Given, $X_t = Z_t - 0.4Z_{t-1} \sim WN(0, \sigma^2)$

$$\begin{aligned} f(\lambda) &= \frac{\sigma^2}{2\pi} [1 + 2\theta \cos \lambda + \theta^2] \\ &= \frac{\sigma^2}{2\pi} [1 + 2 \times (-0.4) \cos \lambda + (-0.4)^2] \\ &= \frac{1}{2\pi} [1 - 0.8 \cos \lambda + 0.16] \end{aligned}$$

AR(1)

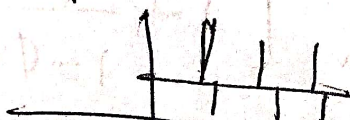
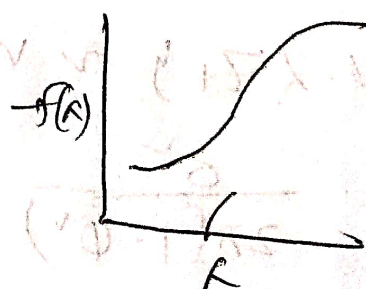
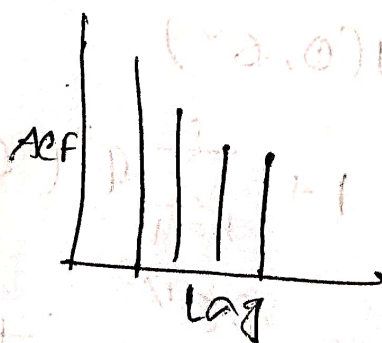
$$X_t = \phi X_{t-1} + Z_t$$

$$\therefore \gamma(h) = \begin{cases} \frac{\sigma^2}{1-\phi^2} & ; h=0 \\ \phi^{|h|} \gamma(0) & ; |h| > 0 \end{cases}$$

$$\begin{aligned} \therefore f(\lambda) &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ih\lambda} |\gamma(h)| \\ &= \frac{\sigma^2}{2\pi} (1 - 2\phi \cos \lambda + \phi^2)^{-1} \end{aligned}$$

ϕ positive

ϕ negative



Periodogram: $\{X_t\}$ is a stationary process with ACF $\gamma(\cdot)$ and the spectral density $f(\cdot)$.

Periodogram $I_n(\cdot)$ is an analogue of $2\pi f(\cdot)$

$X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{C}^n \rightarrow$ vector of complex numbers
Set of all column vectors with complex valued components.

Let, $w_k = \frac{2\pi k}{n}$, k is any integer

$$\frac{n-1}{2} \leq k \leq \frac{n}{2}$$

$$k = -\left[\frac{n-1}{2}\right] \dots \left[\frac{n}{2}\right]$$

F_n , Fourier frequencies associated with sample size n

$F_n \rightarrow$ A subset of $(-\pi, \pi]$

$$X_k = \frac{1}{\sqrt{n}} \begin{bmatrix} e^{i w_k} \\ e^{2i w_k} \\ \vdots \\ e^{n w_k} \end{bmatrix} \quad (n \text{ vectors are introduced})$$

Here, e_i 's are orthogonal as -

$$e_j^* e_k = \begin{cases} 1 & \text{if } j=k \\ 0 & \text{if } j \neq k \end{cases}$$

e_j^* are the row vectors whose k -th component is the complex conjugate of the k -th component of e_j .

e_i 's are the basis for $\mathbb{C}^n$, $x \in \mathbb{C}^n$ and expressed as -

$$x = \sum_k a_k e_k$$

$$a_k = e_k^* x = \frac{1}{\sqrt{n}} \sum_{t=1}^n x_t e^{-it\omega_k}$$

↓
Discrete Fourier transform of the series x_i .

The periodogram of x_t is the function

$$I_n(\lambda) = \frac{1}{n} \left[\sum_{t=1}^n x_t e^{-it\omega_k} \right]^2$$

$$ARMA(p, n)$$

$$x_t - \phi_1 x_{t-1} - \dots - \phi_p x_{t-p} = z_t + \theta_1 z_{t-1} + \dots + \theta_n z_{t-n}$$

Invertible: If $|\theta| < 1$, then $ARMA(1, 1)$ process

is invertible and z_t is expressed in terms of $x_s, s \leq t$.

$$z_t = x_t - (\phi + \theta) \sum_{j=1}^{\infty} (-\theta)^{j-1} x_{t-j}$$

If $|\theta| > 1$, then the $ARMA(1, 1)$ process is non-invertible, and z_t is expressed in terms of $x_s, s \geq t$,

$$z_t = -\phi \theta^{-1} x_t + (\phi + \theta) \sum_{j=1}^{\infty} (-\theta)^{j-1} x_{t+j}$$

Example: An $ARMA(1, 1)$ process consider the

$ARMA(1, 1)$ process x_t satisfying the equations

$$x_t - 0.5x_{t-1} = z_t + 0.4z_{t-1}; \quad \{z_t\} \sim WN(0, \sigma^2)$$

hence, the autoregressive polynomial is -

$$\phi(z) = 1 - 0.5z.$$

The autoregressive polynomial is zero at $z = 2$.

which is located outside the unit circle. Hence, we conclude that from this there exists a unique $ARMA$ process that is also causal.

$$A(p)z: x_t = \phi x_{t-1} + z_t$$

Existence and Uniqueness:

$$\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p \neq 0 \text{ for all } |z| = 1$$

ausality:

$$\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p \neq 0 \text{ for all } |z| \leq 1$$

here, the MA polynomial is —

$$\theta(z) = 1 + \theta_1 z$$

The MA polynomial has a zero at

$$z = -\frac{1}{\theta_1} = -2.5$$

This implies that x_t is invertible.

Invertibility is equivalent to the condition

$$\theta(z) = 1 + \theta_1 z + \dots + \theta_n z^n \neq 0 \text{ for all } |z| \leq 1$$

$$E[x_t, x_{t-1}] = E\left[\sum_{j=1}^{\infty} \phi^{j-1} z_{t-j} \right]$$

$$= \sum_{j=2}^{\infty} \phi^{j-2} z_{t-j}$$

$$= E\left[z_t + z_{t-1}(\phi + \theta) + \sum_{j=2}^{\infty} \phi^{j-2} z_{t-j} \right]$$

$$= z_t + (\phi + \theta) \sum_{j=2}^{\infty} \phi^{j-2} z_{t-j}$$

$$= (\phi + \theta) E[z_{t-1}] + (\phi + \theta) \sum_{j=2}^{\infty} \phi^{j-4} E[z_{t-1}]$$

$$\begin{aligned}
 &= (\phi + \theta) \tilde{\sigma}^2 + (\phi + \theta) \tilde{\sigma}^2 \left[1 + \phi^2 + \phi^4 + \dots + \phi^{2n} \right] \\
 &= (\phi + \theta) \tilde{\sigma}^2 + (\phi + \theta) \tilde{\sigma}^2 (\phi^2) (1 - \phi^2)^{-1} \\
 &= \tilde{\sigma}^2 (\phi + \theta) \left[1 + \frac{(\phi + \theta) \phi^2}{(1 - \phi^2)} \right]
 \end{aligned}$$

An AR(1) process, first order auto regression

$$X_t - \phi X_{t-1} = Z_t$$

An AR(1) model -

$\{X_t\}$ be an AR(1) process with mean μ , defined by the equation -

$$X_t - \mu = \phi (X_{t-1} - \mu) + Z_t \text{ where } |\phi| < 1 \text{ and}$$

$$Z_t \sim WN(0, \tilde{\sigma}^2)$$

$$\sigma^2(n) = \frac{\phi^{1n} \tilde{\sigma}^2}{1 - \phi^2}$$

Distribution of ACF: for inference, we

need the sampling distribution of ACF. The distribution of $\hat{\rho}(h)$ is intractable for samples from even the simplest time series models.

It can usually be well approximated by a normal distribution for large sample sizes.

ARMA:

$$\hat{\rho}_k = (\hat{\rho}(1) \dots \hat{\rho}(k))'$$

$$\hat{\rho} \approx N(\rho, n^{-1}w)$$

w is given by the Bartlett's formula:

$$w_{ij} = \sum_{k=1}^{\infty} [\rho(k+i) + \rho(k-i) - 2\rho(i)\rho(k)] \times [\rho(k+j) + \rho(k-j) - 2\rho(j)\rho(k)]$$

Maximum likelihood estimation:

The Gaussian likelihood for an ARMA process:

$$L(\phi, \theta, \sigma^2) = \frac{1}{\sqrt{(2\pi\sigma^2)^n} \pi_0 \dots \pi_n} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{j=1}^n \frac{(x_j - \hat{x}_j)^2}{\pi_{j-1}} \right\}$$

Differentiating $\ln L(\phi, \theta, \sigma^2)$ partially w.r.t. with respect to σ^2 and noting that $\hat{x}_j$ and π_j are independent of σ^2 , we find that the maximum likelihood estimator $\hat{\phi}$,

maximum likelihood estimators:

$$\hat{\theta}^v = n^{-1} s(\hat{\phi}, \hat{\theta})$$

$$\text{where, } s(\hat{\phi}, \hat{\theta}) = \sum_{j=1}^n (x_j - \hat{x}_j) / \pi_{j-1}$$

and $\hat{\phi}, \hat{\theta}$ are the values of ϕ, θ that minimize

$$l(\phi, \theta) = \ln(n^{-1} s(\phi, \theta)) + n^{-1} \sum_{j=1}^n \ln \pi_{j-1}$$

maximization of $l(\phi, \theta)$ must be done numerically.

Initial values for ϕ and θ can be obtained by ITSM using the methods, the program then searches systematically for the values of ϕ and θ that minimize the reduced likelihood and computes the corresponding maximum likelihood estimate of θ^v .

AIC criterion:

choose $p, \alpha, \phi_p, \theta_\alpha$ to minimize

$$AIC = -2 \ln l(\phi_p, \theta_\alpha, s(\phi_p, \theta_\alpha)/n) + 2(p + \alpha + 1)n / (n - p - \alpha - 2)$$

memory shortening:

$$\tilde{Y}_t = Y_t - \hat{\Phi}(z) Y_{t-2}$$

$$\text{and } \tilde{Y}_t = Y_t - \hat{\Phi}_1 Y_{t-1} - \hat{\Phi}_2 Y_{t-2}$$

with the aid of the five step algorithm described below, we classify $\tilde{Y}_t$ and take one of the following three courses of action.

~~target~~: Declare Y_t to be long memory and form $\tilde{Y}_t$ using $\tilde{Y}_t = Y_t - \hat{\Phi}(z) Y_{t-2}$

m. Declare Y_t to be moderately long-memory and form $\tilde{Y}_t$ using

$$\tilde{Y}_t = Y_t - \hat{\Phi}_1 Y_{t-1} - \hat{\Phi}_2 Y_{t-2}$$

3. Declare Y_t to be short-memory for each $z = 1, 2, \dots, 15$, we find the value $\hat{\Phi}(z)$ of Φ that minimizes

$$ERP(\Phi, z) = \frac{\sum_{t=1}^T [Y_t - \Phi Y_{t-2}]^2}{\sum_{t=2}^T Y_t^2}$$

we then define

$$ERP(z) = ERP(\hat{\Phi}(z), z).$$