Time Series Analysis: An Overview

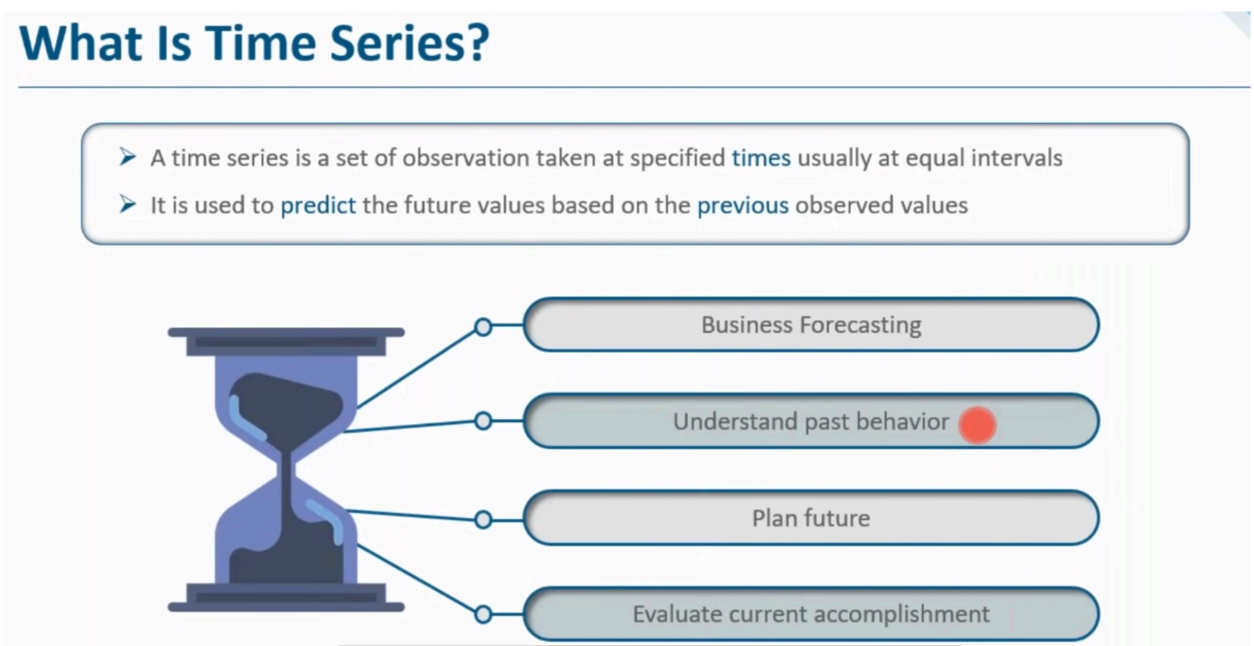
1. Definition of Time Series

A **time series** is a sequence of data points collected or recorded at specific time intervals, often in uniform intervals such as hourly, daily, weekly, monthly, or yearly. Unlike cross-sectional data that captures information at a single point in time, time series data tracks changes over time, providing a dynamic view of trends, patterns, and relationships.

2. Need for Time Series Analysis

Time series analysis is crucial in various fields for the following reasons:

- **Forecasting:** Predicting future trends based on historical data (e.g., stock prices, weather forecasting, sales forecasting).
- **Seasonality Detection:** Identifying regular patterns or cycles over specific periods (e.g., increased sales during holidays).
- **Trend Analysis:** Understanding long-term movements in data, such as upward or downward trends (e.g., population growth).
- **Anomaly Detection:** Identifying unusual events or outliers in the data that may require further investigation (e.g., fraud detection).
- **Causal Analysis:** Exploring relationships between variables over time (e.g., the effect of a marketing campaign on sales).



3. Components of Time Series

Time series data can typically be decomposed into four key components:

- **Trend:** The long-term movement in the data, which can be upward, downward, or flat. Trends represent the overall direction of the data over time.

  -Simple Example: Suppose a town has been established with few unconstructed house, then a person comes and opens up a hardware shop there , for few months his sale was high but after all the constructions has been completed his sales downs, so this is called trend which occurs for long period of time either uptrend or downtrend.

  Stock Market Indexes

- **Example:** Over the long term, stock market indexes like the S&P 500 generally exhibit an upward trend, reflecting overall economic growth and increased corporate profitability. Despite short-term fluctuations and market corrections, the long-term trend is often positive.

  Global Temperature Changes

- **Example:** Data on global temperatures over several decades show a clear upward trend, indicating global warming. This long-term increase is driven by factors such as greenhouse gas emissions and climate change.

  Population Growth

- **Example:** The population of many countries or regions tends to show an upward trend over time due to factors like birth rates exceeding death rates. For instance, the population of India has been steadily increasing for decades.

Company Revenue

- **Example:** A technology company's revenue might exhibit a steady upward trend over several years as the company grows its market share, introduces new products, and expands globally.

Urbanization

- **Example:** The percentage of people living in urban areas versus rural areas has shown an upward trend globally, reflecting the ongoing shift towards urbanization.

- **Seasonality:** Regular patterns or cycles in the data that occur at fixed intervals due to external factors, such as time of year, day of the week, etc.

  Example 1: Retail Sales During the Holiday Season
  Retail sales often show a strong seasonal pattern, with significant increases during the holiday season (e.g., November and December). This pattern repeats every year, driven by consumer behavior related to holidays like Christmas and Black Friday.

  Example 2: Electricity Consumption
  Electricity usage tends to peak during the summer months due to the increased use of air conditioning. This seasonal pattern repeats each year, usually around the same months.

  Example 3: Tourism Trends
  Tourist arrivals in beach destinations typically peak during the summer months (June to August) and dip during the winter months. This pattern is driven by the weather and vacation schedules.

- **Cyclic Patterns:** These are fluctuations in the data with no fixed period but typically occur over a longer duration, often influenced by economic or business cycles.

  Example 1: Business Cycles
  Economic indicators like GDP, unemployment rates, or stock market indices may exhibit cyclic patterns corresponding to periods of expansion and recession. These cycles do not have a fixed duration but generally last several years.

  Example 2: Commodity Prices
  Prices of commodities such as oil, gold, or agricultural products may exhibit cyclic behavior influenced by factors like supply and demand, geopolitical events, and economic conditions. For instance, oil prices may rise during periods of economic growth and fall during recessions.

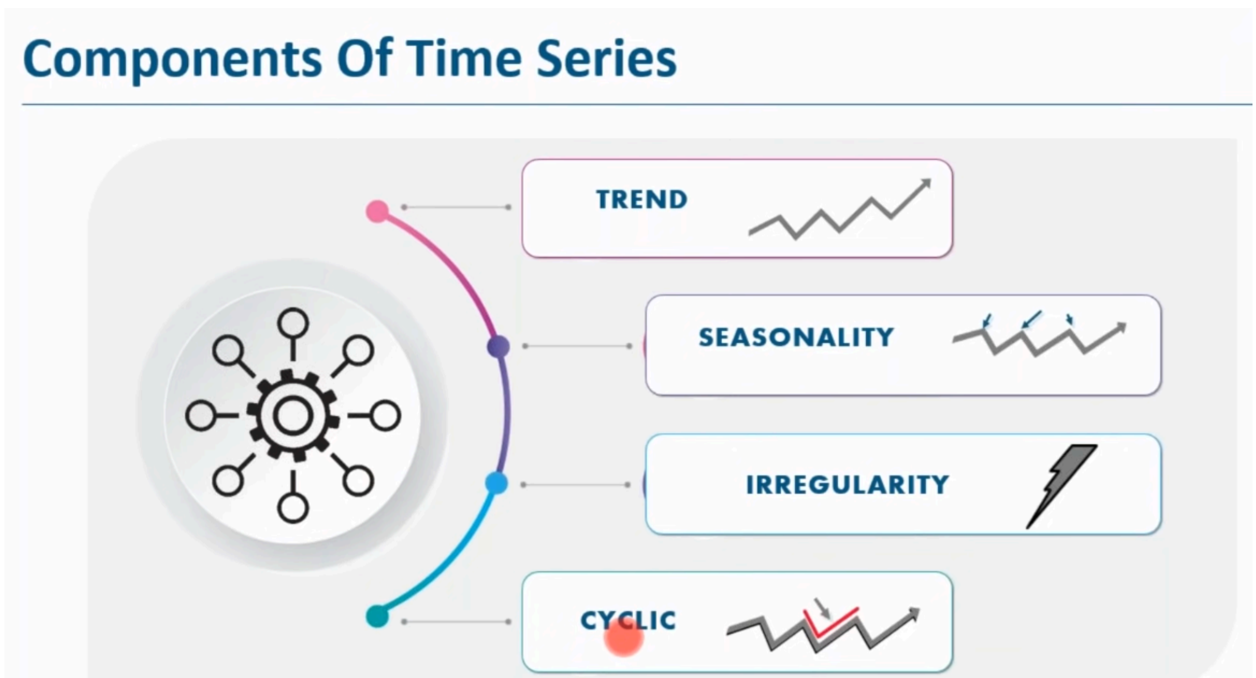
  Example 3: Real Estate Market Trends
  The real estate market often experiences cycles of boom and bust. During a boom, property prices rise as demand increases, and during a bust, prices fall as demand decreases. These cycles can span several years and do not have a fixed period.

- **Irregular or Random Component (Noise):** These are unpredictable variations in the data that do not follow any pattern. They are the residuals after trend, seasonality, and cyclic patterns have been removed.

  Example 1: Stock Market Fluctuations
  Day-to-day fluctuations in stock prices often contain a lot of noise, driven by random events such as breaking news, unexpected economic reports, or sudden changes in investor sentiment. These fluctuations do not follow any predictable pattern.

Example 2: Daily Weather Variations
While there are seasonal trends in weather, daily variations can be highly irregular due to factors like sudden storms, unexpected changes in wind patterns, or other unpredictable atmospheric conditions.

Example 3: Sales Spikes Due to Unexpected Events
A sudden spike in sales of a particular product might occur due to an unexpected event, such as a viral social media campaign or a sudden change in consumer preferences. These spikes are often unpredictable and do not follow any regular pattern.



When to not use Time Series?

1. Lack of Temporal Order

- **Scenario:** The data does not have a natural time component or chronological order.
- **Example:** Data that consists of independent observations without a time sequence, such as survey responses or cross-sectional data, where each data point is collected at a single point in time.

2. Insufficient Data

- **Scenario:** There is not enough historical data to establish patterns or trends.
- **Example:** Time series analysis often requires a substantial amount of historical data to identify trends, seasonality, and cyclic patterns. With too little data, it becomes challenging to make reliable predictions.

3. Non-Stationary Data Without Transformations

- **Scenario:** The data is highly non-stationary, and appropriate transformations or differencing cannot stabilize its mean and variance.
- **Example:** Some time series data may exhibit extreme non-stationarity due to factors like abrupt structural changes, which might require complex modeling beyond standard time series methods.

4. High Noise-to-Signal Ratio

- **Scenario:** The data has a high level of noise compared to the signal, making it difficult to identify meaningful patterns.

- **Example:** If the time series data is dominated by random fluctuations and does not exhibit clear trends or seasonality, time series models may struggle to provide useful forecasts.

5. Complex and Dynamic Relationships

- **Scenario:** The data involves complex relationships between multiple variables that cannot be captured by univariate time series models.
- **Example:** Situations where multiple factors influence the outcome and interactions between these factors are dynamic, such as in systems where external events have significant impact, may require multivariate approaches or causal modeling.

6. Short-Term Focus

- **Scenario:** The focus is on short-term predictions or decisions rather than long-term trends.
- **Example:** For very short-term predictions where immediate responses are required, other techniques like machine learning or real-time analytics might be more suitable than time series analysis.

7. Exogenous Factors Dominating

- **Scenario:** External factors have a dominant influence on the data, overshadowing any internal patterns.
- **Example:** In cases where external shocks or events (e.g., sudden policy changes, natural disasters) have a strong impact on the data, time series models might need to incorporate these factors explicitly or might not perform well without them.

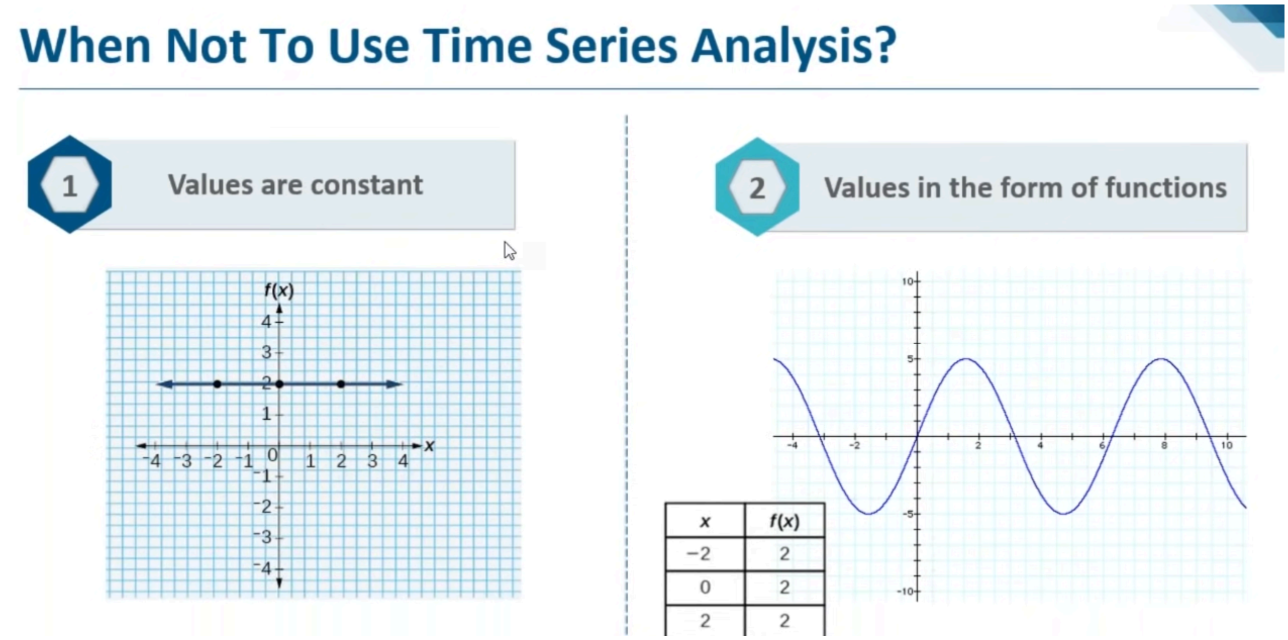
8. Non-Quantitative Data

- **Scenario:** The data is qualitative or categorical rather than numerical.
- **Example:** When dealing with data like text or categorical variables where time series methods are not applicable, other analytical approaches such as text mining or categorical data analysis might be used.

9. Large Seasonal and Cyclic Fluctuations

- **Scenario:** The data exhibits extremely large and unpredictable seasonal or cyclic fluctuations that overshadow any underlying trend.
- **Example:** In cases where seasonality or cyclic patterns are so dominant that they obscure trends, specialized models or adjustments may be needed to properly capture the underlying trends.

In these cases, alternative methods like cross-sectional analysis, machine learning algorithms, or other statistical techniques might be more appropriate for analyzing the data and making predictions.



<u>Case 1: Constant values:-</u>

Suppose you have sales of 500 coffees in your cafe this month and in previous month also you were having sales of 500 coffees and now you have to predict sales of upcoming months, so here time series cannot be applied as present and historical values are both equal or constants.

Stationarity in Time Series

Definition: Stationarity in a time series refers to the property of a time series where its statistical properties, such as mean, variance, and autocorrelation, are constant over time. A stationary time series does not exhibit trends or seasonal effects that change over time, making it easier to model and predict.

Types of Stationarity

1. **Strict Stationarity:**
    - **Definition:** A time series is strictly stationary if its joint distribution remains the same when shifted in time. This means all statistical properties, including higher-order moments, are invariant to time shifts.
    - **Example:** For strictly stationary data, the distribution of values at time ttt and t+kt+kt+k should be the same for all kkk, and not just their means and variances.

2. **Weak Stationarity (or Second-Order Stationarity):**
    - **Definition:** A time series is weakly stationary if its mean and variance are constant over time, and its autocovariance only depends on the time lag between observations, not on the actual time points.
    - **Example:** For weakly stationary data, if you look at the variance and covariance of the series over time, they should be constant and only dependent on the lag between observations.

Importance of Stationarity

- **Modeling:** Many time series models, like ARIMA (AutoRegressive Integrated Moving Average), assume that the data is stationary. Stationary data simplifies the modeling process and improves forecasting accuracy.
- **Consistency:** Stationary series allow for consistent estimation of parameters and better interpretation of model results.
- **Predictability:** Stationary series typically exhibit regular patterns that are easier to predict compared to non-stationary series.

How to Test for Stationarity

1. **Visual Inspection:**
   - Plot the time series data and look for visible trends, seasonal effects, or changes in variance over time. Non-stationary data often shows noticeable patterns or changing statistical properties.
2. **Statistical Tests:**
   - **Augmented Dickey-Fuller (ADF) Test:** Tests the null hypothesis that a unit root is present, indicating non-stationarity. A low p-value suggests rejecting the null hypothesis and indicates stationarity.
   - **Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test:** Tests the null hypothesis that a series is stationary

around a deterministic trend. A high p-value suggests that the data is stationary.

- **Phillips-Perron (PP) Test:** Another test for unit roots, similar to ADF but with adjustments for heteroscedasticity.

How to Achieve Stationarity

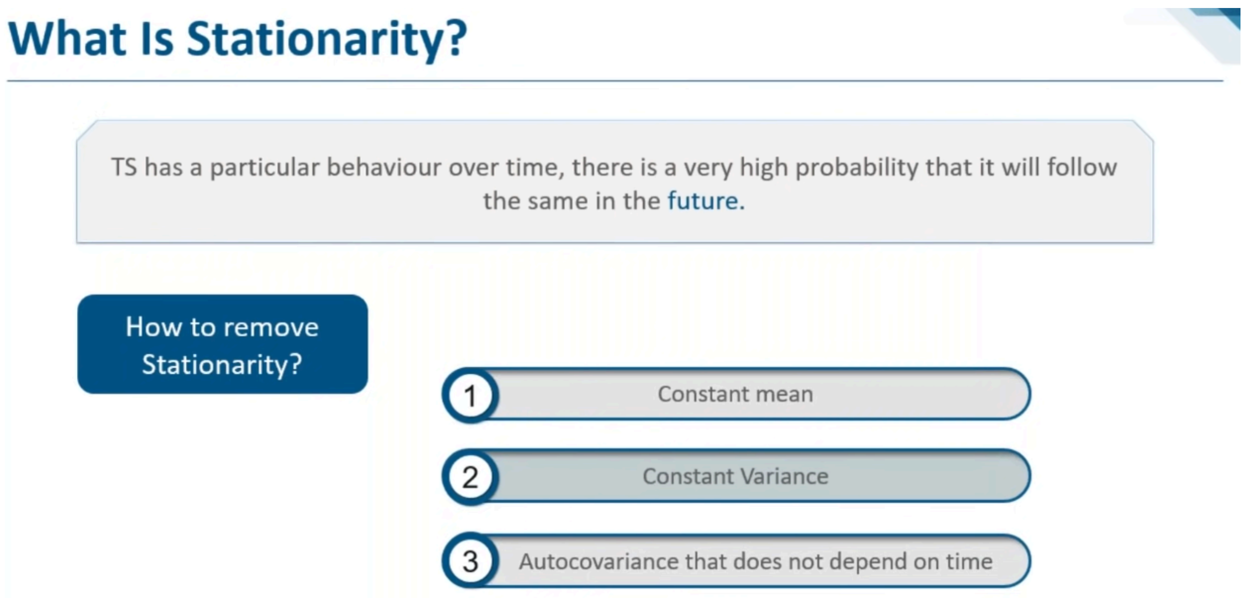
1. **Differencing:**
    - **Definition:** Subtracting the previous observation from the current observation to remove trends and seasonality. This is often applied multiple times if necessary.
    - **Example:** If the original series is yty_tyt, the differenced series is yt−yt−1y_t - y_{t-1}yt−yt−1.
2. **Transformation:**
    - **Definition:** Applying mathematical transformations to stabilize the variance. Common transformations include logarithms, square roots, or Box-Cox transformations.
    - **Example:** Applying a log transformation to a series with exponential growth can help stabilize the variance.
3. **Decomposition:**
    - **Definition:** Separating the time series into its components (trend, seasonality, and residual) and analyzing or modeling each component separately.
    - **Example:** Using seasonal decomposition methods to remove seasonal effects before modeling the residual series.

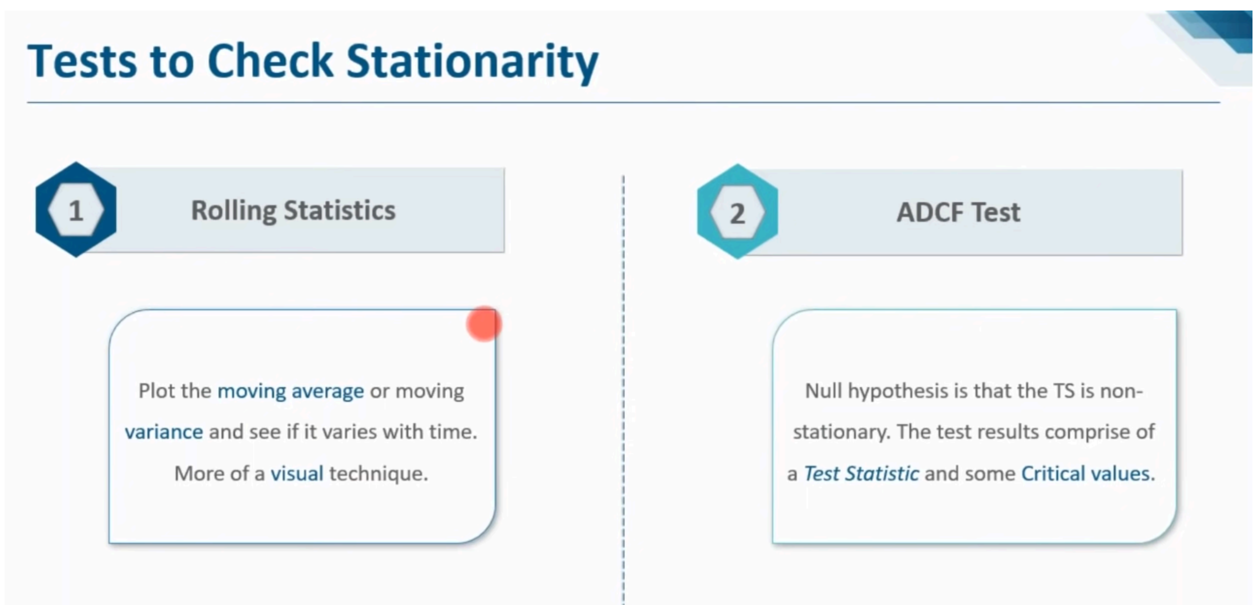
4. **Seasonal Adjustment:**
    - **Definition:** Removing seasonal effects from the data to achieve stationarity.
    - **Example:** Using seasonal decomposition methods like STL (Seasonal and Trend decomposition using Loess) to adjust for seasonal variations.

Example of Non-Stationary vs. Stationary Time Series

- **Non-Stationary Example**:
    - **Original Data:** A time series with an upward trend or seasonal fluctuations, such as monthly sales data with clear seasonal peaks.
- **Stationary Example:**
    - **Transformed Data:** After applying differencing and/or transformations, the time series should exhibit constant mean and variance over time, making it suitable for modeling.

In summary, stationarity is a crucial concept in time series analysis, as it affects the applicability and performance of various modeling techniques. Understanding and achieving stationarity ensures that time series models can produce reliable and accurate forecasts.





<u>ARIMA Model in Time Series</u>

ARIMA stands for **AutoRegressive Integrated Moving Average**. It is a popular and versatile model used for forecasting time series data. The ARIMA model combines three key components: autoregression, differencing (integration), and moving average.

Components of ARIMA

1. **AutoRegressive (AR) Component:**
    - **Definition:** The AR component models the relationship between an observation and a number of lagged observations (previous time points). It captures the effect of past values on the current value.
    - **Notation:** AR(p), where p represents the number of lagged observations used.
    - **Example:** If p=2p = 2p=2, the AR component uses the last two values of the time series to predict the current value.
2. **Integrated (I) Component:**
    - **Definition:** The I component involves differencing the time series data to make it stationary. Differencing is the process of subtracting the previous observation from the current observation to remove trends or seasonality.
    - **Notation:** I(d), where d represents the number of times differencing is applied.
    - **Example:** If d=1d = 1d=1, the series is differenced once; if d=2d = 2d=2, it is differenced twice.
3. **Moving Average (MA) Component:**
    - **Definition:** The MA component models the relationship between an observation and a residual error from a moving average model applied to lagged observations. It captures the effect of past forecast errors on the current value.

- **Notation:** MA(q), where qqq represents the number of lagged forecast errors used.
- **Example:** If q=2q = 2q=2, the MA component uses the last two forecast errors to predict the current value.

ARIMA Model Notation

The ARIMA model is generally denoted as **ARIMA(p, d, q)**, where:

- p = Number of lag observations in the autoregressive part.
- d = Number of times differencing is applied.
- q = Number of lagged forecast errors in the moving average part.

Steps to Build an ARIMA Model

1. **Visualize and Preprocess the Data:**
    - Plot the time series to identify trends, seasonality, and potential non-stationarity.
    - Apply transformations (e.g., log) if needed to stabilize variance.
2. **Check for Stationarity:**
    - Use statistical tests like Augmented Dickey-Fuller (ADF) or KPSS tests to check if the data is stationary.
    - If non-stationary, apply differencing to make the data stationary.

3. **Identify the AR and MA Terms:**
   - Use plots like the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) to determine the appropriate values for ppp and qqq.
   - **ACF:** Helps to determine the MA(q) term by showing the correlation between the time series and its lagged values.
   - **PACF:** Helps to determine the AR(p) term by showing the correlation between the time series and its lagged values after removing the effect of intermediate lags.
4. **Fit the ARIMA Model:**
   - Use statistical software or libraries (e.g., `statsmodels` in Python) to fit the ARIMA(p, d, q) model to the data.
   - Estimate the parameters of the AR, I, and MA components.

5. **Diagnose and Validate the Model:**
   - Check the residuals (errors) of the model to ensure they resemble white noise (i.e., they should be randomly distributed with no patterns).
   - Use model evaluation metrics such as AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) to compare different models and select the best one.
6. **Make Forecasts:**
   - Use the fitted ARIMA model to generate forecasts for future time periods.
   - Evaluate the forecast accuracy using metrics like Mean Absolute Error (MAE) or Mean Squared Error (MSE).

Example of ARIMA Model

Suppose you have monthly sales data with a trend and seasonal component. To model this data using ARIMA:

1. **Preprocess:** Transform the data (e.g., take the log) to stabilize variance if needed.
2. **Check Stationarity:** Apply differencing (e.g., d = 1) if the data is not stationary.
3. **Identify Terms:** Plot ACF and PACF to determine values for p and q.
4. **Fit Model:** Suppose you find ARIMA(2, 1, 1) to be a good fit based on your analysis.
5. **Diagnose:** Check residuals for white noise.
6. **Forecast:** Use the fitted ARIMA(2, 1, 1) model to forecast future sales.

Advantages of ARIMA

- **Flexibility:** Can model a wide range of time series patterns by adjusting p, d, and q.
- **No Need for Seasonality:** While ARIMA itself does not model seasonality, it can be extended to SARIMA (Seasonal ARIMA) to handle seasonal effects.

Disadvantages of ARIMA

- **Complexity:** Selecting the right parameters can be complex and time-consuming.
- **Requires Stationarity:** The data must be made stationary, which may not always be straightforward.

- **Limited Handling of Seasonality:** Standard ARIMA does not handle seasonality; SARIMA or other models are needed for seasonal data.

ARIMA is a powerful tool for forecasting and analyzing time series data, provided the underlying assumptions and requirements are met.