Time Series Content

Reading Material

PW SKILLS

Topics:

1. What Is A Time Series?
- Definition and Significance
- Components of Time Series
- Applications of Time Series Analysis

2. ARIMA (AutoRegressive Integrated Moving Average)
- Introduction to ARIMA
- Understanding ARIMA Model
- Steps to Build an ARIMA Model

3. ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function)
- Introduction to ACF
- Introduction to PACF
- Applications of ACF and PACF

4. Time-Dependent Seasonal Components
- Understanding Seasonality in Time Series
- Modeling Seasonal Components
- Applications of Seasonal Analysis

5. Autoregressive (AR) Models
- Introduction to AR Models
- Mathematical Formulation
- Applications of AR Models

6. Moving Average (MA) Models
- Introduction to MA Models
- Mathematical Formulations
- Applications of ARMA Models

1.What is a time Series?

A time series consists of data points gathered or documented at regular time intervals in chronological order. It shows the values of specific variables or a group of variables throughout a period, enabling the examination of trends, patterns, and other time-related behaviors. Time series data is frequently utilized in a range of fields including economics, finance, environmental science, and others.

Example Data:
- Stock Price prediction
- Yearly Sales prediction etc

A sequence of observations taken one after the other throughout time is called a time series. Time series data records changes over time, in contrast to cross-sectional data, which presents information at a single point in time.

Time series analysis is useful in determining underlying trends, such as growth or decline, in data over time.

- **Forecasting:** Time series models are crucial for creating future value predictions based on previous data, which is important for making decisions in marketing, finance, and other areas.
- **Seasonality and anomaly detection:** Recognising regularities helps companies make appropriate plans, including allocating inventory according to demand cycles.Time series analysis is useful for identifying anomalies, or outliers, which may point to potential issues or opportunities.
- **Control and Optimisation:** Processes can be made more efficient by making adjustments to production schedules in manufacturing, for example, by ongoing monitoring and analysis.

Components of Time Series

Trend: A sustained rise or fall in the data. It shows the data's overall direction across a longer time frame.

Example: The slow increase in global temperatures over several decades is one example.

Seasonality: Recurring behavioral cycles or patterns across a predetermined time frame, like a day, a month, or a year.

Example: A rise in sales throughout the Christmas season.

Cyclic Patterns: Data fluctuations that happen at odd intervals and are frequently connected to external variables such as economic cycles are known as cyclical patterns.

Example: A business cycle of expansion and contraction is an example.

Noise: Unpredictable changes in the data that cannot be linked to cycles, seasonality, or trends are known as noise (or irregular component). These are erratic and typically the result of unanticipated circumstances.

Example: An unexpected increase in demand brought on by a social media fad gone viral.

Applications of Time Series Analysis

Economics and Finance: Stock price prediction,Inflation rate analysis,Forecasting sales.

Weather Forecasting: Predicting rainfall,Temperature forecasting,Monitoring climate change.

Supply Chain Management: Demand forecasting,Inventory management,Transportation planning.

Healthcare: Monitoring patient vital signs over time,Predicting the spread of diseases,Analyzing healthcare trends.

Energy Sector: Predicting electricity consumption,Renewable energy production forecasting,Analyzing energy market trends.

Social Media Analysis: Monitoring sentiment trends,Analyzing engagement patterns over time,Predicting viral trends.

ARIMA (AutoRegressive Integrated Moving Average):

AutoRegressive Integrated Moving Average is abbreviated as ARIMA. A lagged forecast error (MA), the difference between prior values (I), and the time series itself (AR) are the three factors that this class of models uses to describe a particular time series. **Finding the underlying structure in time series data and using it for forecasting is the main objective of ARIMA.**
ARIMA is one of the most popular and widely used statistical methods for time series forecasting. ARIMA models are particularly effective when dealing with non-stationary time series data.

Key Characteristics:

AR(AutoRegression): A model known as AR (AutoRegression) makes advantage of the relationship between one observation and several lagged observations.

I (Integrated): The process of making the time series stationary by differencing the raw observations (i.e., to remove trends or seasonality).

MA (Moving Average): A model that applies a moving average model's residual error to lagged observations, utilizing the relationship between an observation and the error.

Understanding ARIMA Model

Typically, the ARIMA model is written as ARIMA(p, d, q), where:

p: The number of lag observations (i.e., the number of previous values to be used for predicting the future value) contained in the model is denoted by the letter p (AutoRegression order).

d: The number of times the data must be differenced to become stationary is indicated by the letter d (integrated order).

q: The size of the moving average window, or the total number of historical forecast errors incorporated into the model, is denoted by q (moving average order).

For instance, to anticipate future sales using monthly sales data, we would create an ARIMA model by choosing suitable values for p, d, and q based on the properties of the data.

Steps to Build an ARIMA Model:

1. Check for Stationarity: Make sure the time series data is stationary, meaning that its mean, variance, and autocorrelation structure remain constant over time, before constructing an ARIMA model.
Visual inspection, statistical tests such as the Augmented Dickey-Fuller (ADF) test, or plotting autocorrelation functions (ACF) can all be used to determine stationarity.

2. Differencing (if necessary): If the time series data is not steady, use differencing to eliminate seasonality and trends.To do differencing, subtract the previous observation from the present observation .Ex:(Y(t) - Y(t-1)).

3. Identify Model Parameters (p, d, q): To determine the proper values for p and q, use plots of autocorrelation (ACF) and partial autocorrelation (PACF).The number of times the data was differenced to ensure stationarity determines the number of differences (d).

4. Fit the ARIMA Model: Using statistical tools (such as Python's statsmodels package), we can fit the ARIMA model to the data once the p, d, and q values are identified.
To make sure the model is a good fit, examine model diagnostics (such as residual plots) and later use the fitted model to provide forecasts.

5. Model Diagnostics: Once the model has been fitted, make sure the residuals, or errors, have a mean of zero and no autocorrelation, similar to that of white noise.
To verify the model's assumptions, use methods such as residual plots..

6. Forecasting: Use the fitted ARIMA model to make future predictions. The accuracy of the forecasts can be evaluated using metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), etc.

ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function):

ACF and PACF are essential tools in time series analysis, particularly in identifying the structure of a time series and determining the appropriate parameters for models like ARIMA. They help analyze the dependence between observations at different time lags and are widely used in various time series forecasting applications.

Introduction to ACF (Autocorrelation Function):

Definition: The correlation between a time series and its lag values is measured by the Autocorrelation Function (ACF). Put more simply, it illustrates the relationship between a data point at one time and other data points at various time intervals (lags).

In mathematical terms, the correlation between the series values at time t and time t-k is the autocorrelation at lag k.

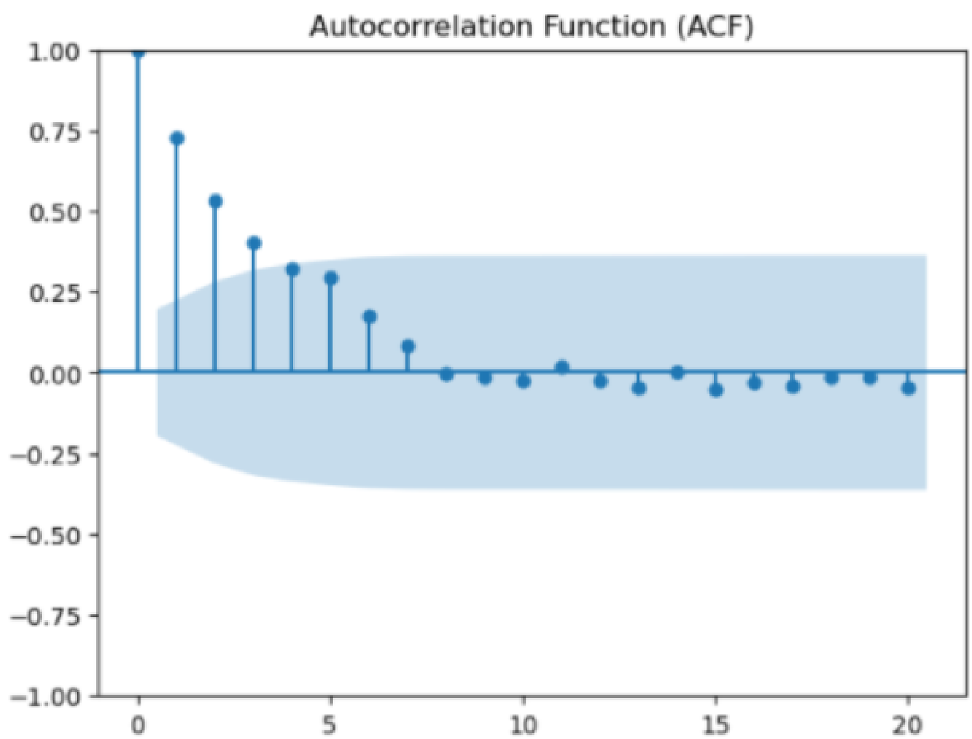
Features:

Range: The ACF values are in the range of -1 to 1.

Interpretation: A positive autocorrelation suggests that a similar-sized data point is likely to come after a given data point.
A data point with a negative autocorrelation suggests that a data point with the opposite magnitude is probably going to come after it.
An autocorrelation around zero indicates that there isn't a linear connection between the data points at that lag.

Example:
If the ACF at lag 1 is 0.8, this means the data point at time t is strongly positively correlated with the data point at time t-1.

Plot:
Example demonstrating a ACF plot on a synthetically generated time series data
An ACF plot displays the autocorrelation coefficients for different lags, helping identify the presence of any repeating patterns or seasonality in the time series.



Introduction to PACF - Partial Autocorrelation Function

Definition: On the other hand, the PACF measures the relationship between a time series and its lagged values, but does so after the influence of intermediate lags has been removed. What do we mean by this? Well, basically speaking, the partial autocorrelation function isolates the direct relationship between a value at some point in time, t, and its lagged value at another point in time, t - k, making sure not to count any relationships due to all the values that fall in between.

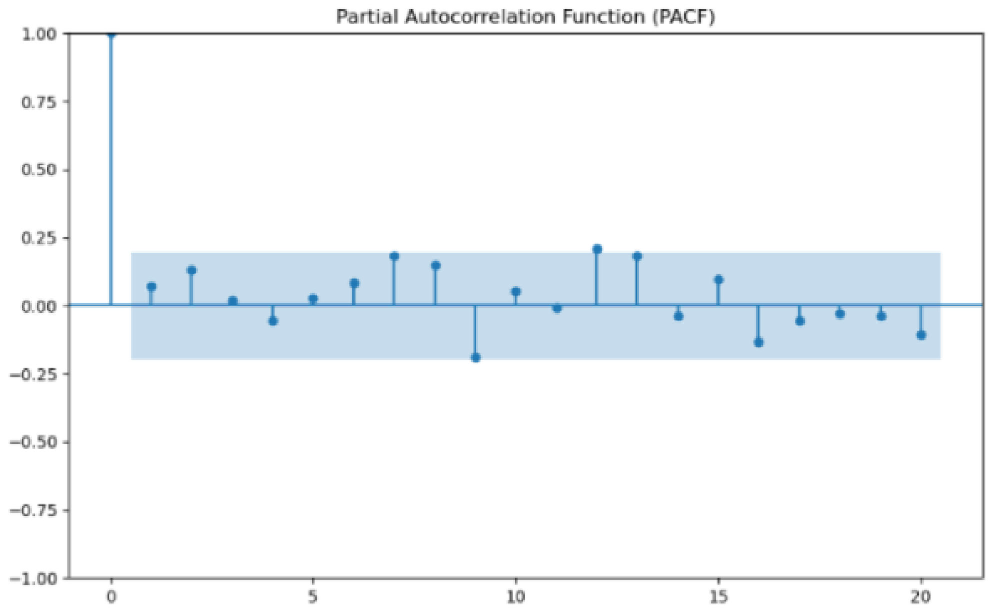
Mathematically:
PACF at lag k is the partial correlation between the series values at time t and time t-k, conditional on all the lags up to k.

Characteristics:

Interpretation:
The PACF at lag k provides the pure correlation between the value at time t and t-k, without accounting for the effects of other lags.
Like ACF, the range of values for PACF are between -1 and 1.
Use in Model Building:
In an AR(p) model, PACF will display significant lags up to p and then fall off, which helps in checking for the correct number of autoregressive terms.

Example:
For instance, this would indicate that if the PACF at lag 2 is 0.5, after accounting for the correlation due to lag 1, data point at time t is still left with a fair positive correlation with data point at time t - 2.

Plot:
Example demonstrating a PACF plot on a synthetically generated time series data
A PACF plot shows partial autocorrelation coefficients against various lags to resist the identification of significant lags, which should be included in the model.



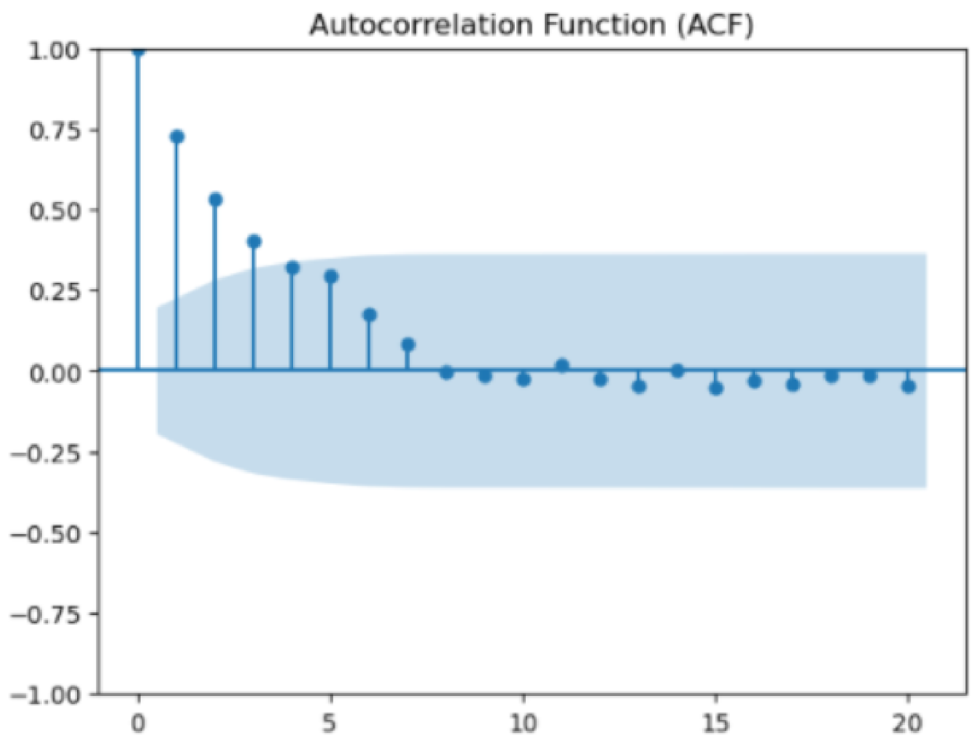
Features:

Range: The ACF values are in the range of -1 to 1.

Interpretation: A positive autocorrelation suggests that a similar-sized data point is likely to come after a given data point.
A data point with a negative autocorrelation suggests that a data point with the opposite magnitude is probably going to come after it.
An autocorrelation around zero indicates that there isn't a linear connection between the data points at that lag.

Example:
If the ACF at lag 1 is 0.8, this means the data point at time t is strongly positively correlated with the data point at time t-1**.**

Plot:
Example demonstrating a ACF plot on a synthetically generated time series data
An ACF plot displays the autocorrelation coefficients for different lags, helping identify the presence of any repeating patterns or seasonality in the time series.



ACF and PACF Applications

- **Identifying Model Parameters**:ACF and PACF plots are crucial in determining what order of AR and MA components exist in an ARIMA model.
- **ACF for MA Models:** The ACF plot in an MA(q) model will have significant spikes up to a lag of q and then drop off.
- **PACF for AR Models:**This will imply that the PACF plot for an AR(p) will have significant spikes up to lag p and then fall away.
- **Finding Seasonality:**ACF plots can be helpful in finding the seasonality in time series data. If the autocorrelation is high at fixed periodic intervals, say every 12 months for monthly data, it indicates seasonality.
- **Residuals Diagnosis:** One could also use ACF and PACF plots of residuals after fitting a time series model to see if there is any residual autocorrelation. The residuals must ideally appear as white noise with no significant spikes in either the ACF or PACF plot.
- **Understanding Relationships in Data:** ACF and PACF help in understanding the magnitude of dependency of data points on their previous values, which is a key in the selection of appropriate forecasting methods.
- Anomaly detection From these ACF and PACF plots, the analyst can observe any anomalies or unnatural variations within the axial series, such as a sudden shift or break in pattern.
- **Lag Order Determination in Models:** The ACF and PACF plots provide recommendations about the inclusion of the number of lags, or p, to include in vector autoregressive-or VAR-and other time series models.

Time-Dependent Seasonal Components

Understanding Seasonality in Time Series

Definition of Seasonality: Seasonality refers to periodic fluctuations in a time series that occur at regular intervals due to seasonal factors. These fluctuations are typically driven by calendar-related effects, such as seasons, holidays, or specific times of the year. For instance, retail sales often increase during the holiday season, and temperature data shows regular patterns throughout the year.

Characteristics of Seasonal Patterns:

- **Periodicity:** Seasonality is characterized by regular intervals, such as monthly, quarterly, or yearly cycles.
- **Repetition:** The patterns repeat at consistent intervals.
- **Magnitude:** The strength of seasonal effects may vary in different periods.

Identifying Seasonality: Seasonality can be identified through visual inspection of time series plots, statistical tests, and decomposition methods. Tools like autocorrelation plots and periodograms can also reveal periodic patterns.

Modeling Seasonal Components

Seasonal Decomposition: Seasonal decomposition involves breaking down a time series into its components:

- **Trend:** Long-term movement in the data.
- **Seasonality:** Repeating patterns at fixed intervals.
- **Residuals:** Irregular or noise component.

Decomposition Methods:

- **Additive Decomposition:** Suitable when seasonal fluctuations are constant in magnitude. $Y(t)=T(t)+S(t)+R(t)$
- **Multiplicative Decomposition:** Suitable when seasonal fluctuations vary proportionally with the level of the time series. $Y(t)=T(t)\times S(t)\times R(t)$

Libraries for Decomposition:

- **statsmodels:** seasonal_decompose function for additive and multiplicative models.
- **pmdarima:** auto_arima for automated model selection, including seasonal components.

Seasonal Models:

- **SARIMA (Seasonal ARIMA):** Extends ARIMA to handle seasonality by adding seasonal terms. ARIMA(p,d,q) with seasonal terms (P,D,Q,s)
- **Exponential Smoothing State Space Model (ETS):** Models seasonality through smoothing techniques.
- **Seasonal Neural Networks:** Deep learning models designed to capture complex seasonal patterns.

Handling Seasonality in Machine Learning:

- **Feature Engineering:** Create features that capture seasonal effects (e.g., month of the year, day of the week).
- **Time Series Cross-Validation:** Ensure that models are tested on different seasonal periods to validate performance.

Applications of Seasonal Analysis

1. Retail and Sales Forecasting:
- Predicting inventory needs based on seasonal demand patterns.
- Optimizing marketing strategies and promotions during peak seasons.

2. Energy Consumption:
- Forecasting energy needs based on seasonal temperature variations.
- Planning for peak load periods and managing resources efficiently.

3. Finance and Economics:
- Analyzing seasonal trends in stock prices or economic indicators.
- Adjusting financial forecasts for recurring seasonal effects.

4. Agriculture:
- Planning for crop planting and harvesting based on seasonal weather patterns.
- Forecasting yield and managing supply chains.

5. Tourism:
- Predicting tourist arrivals and planning resources for peak seasons.
- Developing marketing campaigns aligned with seasonal travel trends.

Autoregressive (AR) Models

Introduction to AR Models

The autoregressive model is, in general, a time series model relating the past to the future. The general idea is that the value of a variable with respect to time depends linearly on preceding values. Autoregressive models have seen widespread use in areas like finance, economics, and signal processing, among others, mainly because of their simplicity and modeling of time dependencies.

Key Points:

Model Concept: The value at a time t is modeled as a linear function of its past values plus some noise.
Notation: An AR model of order p is denoted as AR(p), where p represents the number of lagged observations used.

Mathematical Formulation:

The general form of an Autoregressive model of order ppp (AR(p)) can be expressed as:
$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \varepsilon_t$

where:

- X_t is the value of the time series at time t.
- $\phi_1, \phi_2, \ldots, \phi_p$ are the parameters of the model.
- ε_t is the white noise error term at time t, assumed to be normally distributed with zero mean and constant variance.

Key Parameters:

- **Lag Order p:** The number of past values used in the model.
- **Coefficients ϕ_i:** These determine the influence of each lagged value on the current value.

Applications of the AR models

Autoregressive models can be used in the following applications:

Economic Forecasting: These are used to predict economic indicators including GDP, inflation rates, and unemployment by using AR models. They allow insight into past economic behavior for the purpose of forecasting future trends.

Financial Markets: Financially, AR models are employed in predicting stock prices, modeling volatility, and managing risks. They enable the analysis of market trends and making informed investment decisions accordingly.

Signal Processing: The models in AR find prominent applications in digital signal processing for analyzing and predicting signals, which form a part of speech processing, image reconstruction, and noise reduction.

Weather Forecast: These models find their applications in predicting weather patterns, examining historical data on temperature, precipitation, and other meteorological phenomena.

Example Use Case: Suppose an AR model is used to predict daily sales data for a retail store based on past sales. The business will, therefore, identify various patterns and trends in its historical sales, which could then be factored into decisions about inventory and marketing strategies.

Moving Average (MA) Models

Introduction to MA Models

The MA models are classes of time series models in which the time series is presented as a function of past white noises or error terms. As opposed to AR models, which model the current value based on past values of the series, MA models take into consideration forecast errors from previous periods. MA models are often used along with AR models to form the so-called ARMA models, which can capture both autoregressive and moving average components.

Mathematical Formulations:

An MA model of order qqq, denoted as MA(q), can be mathematically expressed as:

$X_t=\mu+\epsilon_t+\theta_1\epsilon_{t-1}+\theta_2\epsilon_{t-2}+\cdots+\theta_q\epsilon_{t-q}$

Where:

- X_t is the value of the time series at time t,
- μ is the mean of the series,
- ϵ_t represents the error term (or white noise) at time t,
- $\theta_1,\theta_2,\ldots,\theta_q$ are the model parameters that determine the influence of past errors on the current value.

Applications of ARMA Models:

ARMA models are frequently used to analyze and forecast time series data across a variety of fields. They do this by combining both AR (AutoRegressive) and MA (Moving Average) components. Typical uses for them include:

Financial Time Series: By capturing both short- and long-term trends, ARMA models are widely used to forecast financial measures such as stock prices, interest rates, and exchange rates.

Economic Data Analysis: ARMA models provide insights into future economic conditions by forecasting macroeconomic indices including GDP growth, inflation, and unemployment rates.

Weather Forecasting: By modeling variables such as temperature, precipitation, and wind speed, ARMA models can predict weather patterns by taking into account both current changes and longer-term cycles.

Signal Processing: To reduce noise and do other signal processing tasks, engineers employ ARMA models.