**Assignment: 04**

**Q1. What do you mean by ‘Time Series’ and ‘Time Series Analysis’?**

* **Time Series:** Time series can be regarded as a set of values {xt }, which represents measurement taken at different (sequential) time periods, t=1,2,3,…,n.
* Time itself is a clearly a continuous variable, but measurement in many cases are made at specific point of time , so appear as discrete observation.
* Typically, a single variable or value is analyzed at each point in time even if multiple variables are recorded for that point.
* Examples: Weather data, Earthquake Monitoring, Rainfall measurements, Temperature readings, Heart rate monitoring (EKG), Brain monitoring (EEG), Quarterly sales, Stock prices, Automated stock trading, Industry forecasts.
* **Time Series Analysis:** Data often measured or defined for times that have equal intervals between them E.g.:
* Temperature Data of Every 5 minutes.
* Close of trading every day.
* Time series analysis can be Univariate data, Bivariate or Multivariate
* Time series almost always exhibit some degree of autocorrelation hence Data analysis can be done to identify pattern using autocorrelation function / correlogram.

**Q2. List out Importance of TSA.**

* **Importance of TSA:** It is useful by many organizations to forecast their business profit or loss trends. Thus, important business decisions can facilitate development.
* It is useful to compare the present trend with the past trend that has already happened so the future trend can be estimated and prepared.
* The cycle variations over a period using time series will allow us to understand the business cycle quite effectively.
* It is useful in the quality control process to predicate the quality trend over time.
* If helps to understand the complex signal pattern It is useful to understand how an event can change its feature over a period of time.
* With TSA, the reliability, flexibility, and other important features can be predicated.

**Q3. With suitable plots, describe Components of TSA?**

* **Components of TSA:**

1. **Trend:** A long-term upward or downward movement in the data, indicating a general increase or decrease over time.
2. **Seasonality:** A repeating pattern in the data that occurs at regular intervals, such as daily, weekly, monthly, or yearly.
3. **Cycle:** A pattern in the data that repeats itself after a specific number of observations, which is not necessarily related to seasonality.
4. **Irregularity:** Random fluctuations in the data that cannot be easily explained by trend, seasonality, or cycle.
5. **Autocorrelation:** The correlation between an observation and a previous observation in the same time series.
6. **Outliers:** Extreme observations that are significantly different from the other observations in the data.
7. **Noise:** Unpredictable and random variations in the data.

|  |
| --- |
| How To Isolate Trend, Seasonality And Noise From A Time Series – Time  Series Analysis, Regression, and Forecasting |

* **REFERENCES:** <https://www.geeksforgeeks.org/components-of-time-series-data/>

**Q4. With Suitable Mathematical Derivation, Advantages, Disadvantages and Applications Explain:**

* 1. **AR model:** The time period at t is impacted by the observation at various slots t-1, t-2, t-3, ….., t-k. The impact of previous time spots is decided by the coefficient factor at that particular period of time. The price of a share of any particular company X may depend on all the previous share prices in the time series. This kind of model calculates the regression of past time series and calculates the present or future values in the series in know as Auto Regression (AR) model.

**Yt = β₁\* y-₁ + β₂\* yₜ-₂ + β₃ \* yₜ-₃ + ………… + βₖ \* yₜ-ₖ**

* **Advantages of the AR model:**
* **Simplicity:** The AR model is relatively straightforward and easy to understand compared to more complex time series models.
* **Interpretability:** The autoregressive coefficients (φ\_1, φ\_2, ..., φ\_p) provide insights into the relationships and dependencies between the current and past values of the variable.
* **Flexibility:** The AR model can capture different patterns in time series data, including trends, cyclicality, and short-term dependencies.
* **Efficient computation:** Estimating the parameters of the AR model can be computationally efficient, particularly for lower-order models.
* **Disadvantages of the AR model:**
* **Limited modeling capability:** The AR model assumes a linear relationship between the current and past values. It may not capture complex nonlinear dependencies or long-term dependencies.
* **Sensitivity to outliers:** The AR model can be sensitive to outliers or extreme values in the data, which can impact its forecasting performance.
* **Stationarity assumption:** The AR model requires the time series to be stationary for accurate parameter estimation and reliable forecasts.
* **Determining the order:** Selecting the appropriate order (p) of the AR model can be challenging and requires statistical techniques or domain knowledge.
* **Applications of the AR model:**
* **Economic forecasting:** AR models are widely used for forecasting economic indicators such as stock prices, exchange rates, and GDP growth.
* **Demand forecasting:** AR models can be applied to predict product demand in various industries, such as retail, manufacturing, and logistics.
* **Weather prediction:** Autoregressive models are utilized in meteorology to forecast weather variables, such as temperature, rainfall, and wind speed.
* **Signal processing:** AR models find applications in signal processing tasks like speech recognition, image processing, and time series analysis of sensor data.

Remember that the AR model is just one type of time series model, and its suitability depends on the specific characteristics and requirements of the data.

* **REFERENCES:** <https://towardsdatascience.com/time-series-models-d9266f8ac7b0>
  1. **MA model:** The time period at t is impacted by the unexpected external factors at various slots **t-1, t-2, t-3, ….., t-k.** These unexpected impacts are known as Errors or Residuals. The impact of previous time spots is decided by the coefficient factor α at that particular period of time. The price of a share of any particular company X may depend on some company merger that happened overnight or maybe the company resulted in shutdown due to bankruptcy. This kind of model calculates the residuals or errors of past time series and calculates the present or future values in the series in know as Moving Average (MA) model.

**Yt = α₁\* Ɛₜ-₁ + α₂ \* Ɛₜ-₂ + α₃ \* Ɛₜ-₃ + ………… + αₖ \* Ɛₜ-ₖ**

* **Advantages of the MA model:**
* **Flexibility:** The MA model can capture short-term dependencies and fluctuations in the time series data, making it suitable for forecasting in scenarios where recent observations carry more weight.
* **Stationarity:** Unlike the autoregressive (AR) model, the MA model does not require the time series to be stationary. It can handle non-stationary data and still produce reliable forecasts.
* **Interpretability:** The moving average coefficients (θ\_1, θ\_2, ..., θ\_q) provide insights into the impact of previous error terms on the current value, allowing for interpretation of the model's behavior.
* **Disadvantages of the MA model:**
* **Limited modeling capability**: The MA model assumes a linear relationship between the current value and the error terms. It may not capture complex nonlinear dependencies or long-term patterns in the data.
* **Identifiability issues:** Estimating the parameters of the MA model can be challenging due to identifiability issues, particularly when the lagged error terms are highly correlated.
* **Prediction horizon:** The MA model tends to perform well for short-term forecasting but may struggle with long-term predictions as the impact of past error terms diminishes over time.
* **Sensitivity to outliers:** Like the AR model, the MA model can be sensitive to outliers or extreme values in the data, which can impact its forecasting accuracy.
* **Applications of the MA model:**
* **Finance:** The MA model is widely used in financial time series analysis for modeling and forecasting asset prices, volatility, and trading signals.
* **Quality control:** MA models are applied in quality control processes to detect and forecast deviations or anomalies in manufacturing or production processes.
* **Inventory management:** The MA model can be employed to forecast demand patterns and optimize inventory levels in supply chain management.
* **Network traffic analysis:** MA models find applications in predicting network traffic loads and optimizing network capacity in telecommunications and computer networks.
* **REFERENCES:** <https://towardsdatascience.com/time-series-models-d9266f8ac7b0>
  1. **ARMA model:** This is a model that is combined from the AR and MA models. In this model, the impact of previous lags along with the residuals is considered for forecasting the future values of the time series. Here β represents the coefficients of the AR model and α represents the coefficients of the MA model.

**Yt = β₁\* yₜ-₁ + α₁\* Ɛₜ-₁ + β₂\* yₜ-₂ + α₂ \* Ɛₜ-₂ + β₃ \* yₜ-₃ + α₃ \* Ɛₜ-₃ +………… + βₖ \* yₜ-ₖ + αₖ \* Ɛₜ-ₖ**

* **Advantages of the ARMA model:**
* **Flexibility:** The ARMA model can capture both short-term and long-term dependencies in the data, combining the strengths of autoregressive and moving average models.
* **Adaptability:** The ARMA model can handle time series data with different characteristics, including trends, seasonality, and irregular fluctuations.
* **Stationarity:** The ARMA model can handle non-stationary data by incorporating differencing or integration techniques, such as in the autoregressive integrated moving average (ARIMA) model.
* **Interpretable coefficients:** The autoregressive and moving average coefficients provide insights into the impact of past values and error terms on the current value, enabling interpretation of the model's behavior.
* **Disadvantages of the ARMA model:**
* **Complexity:** The ARMA model can be more complex than individual AR or MA models, especially when higher-order models (p or q) are required.
* **Parameter estimation:** Estimating the parameters (φ\_1, φ\_2, ..., φ\_p, θ\_1, θ\_2, ..., θ\_q) of the ARMA model can be computationally intensive and may require iterative optimization algorithms.
* **Determining the model order:** Selecting the appropriate orders (p and q) of the ARMA model can be challenging and often requires statistical techniques, model selection criteria, or domain knowledge.
* **Sensitivity to outliers:** The ARMA model, like its AR and MA components, can be sensitive to outliers or extreme values in the data, potentially affecting its forecasting performance.
* **Applications of the ARMA model:**
* **Financial forecasting:** The ARMA model is widely used in finance for modeling and forecasting asset prices, stock returns, and volatility.
* **Demand forecasting:** ARMA models are applied in various industries to predict demand patterns for products or services, optimizing inventory management and resource allocation.
* **Climate modeling:** ARMA models find applications in climate and environmental sciences to analyze and forecast temperature, rainfall, and other meteorological variables.
* **Economic analysis:** ARMA models are utilized in economic research to analyze and forecast macroeconomic variables, such as GDP.
* **REFERENCES:** <https://towardsdatascience.com/time-series-models-d9266f8ac7b0>
  1. **ARIMA model:** The ARIMA model is quite similar to the ARMA model other than the fact that it includes one more factor known as Integrated( I ) i.e. differencing which stands for I in the ARIMA model. So in short ARIMA model is a combination of a number of differences already applied on the model in order to make it stationary, the number of previous lags along with residuals errors in order to forecast future values.
* **Advantages of the ARIMA model:**
* **Flexibility:** The ARIMA model can handle time series data with various characteristics, including trends, seasonality, and irregular fluctuations.
* **Stationarity:** The differencing component (d) allows the ARIMA model to handle non-stationary data by transforming it into a stationary series.
* **Comprehensive modeling:** The AR, I, and MA components capture different dependencies and patterns in the data, providing a comprehensive framework for time series analysis and forecasting.
* **Interpretable coefficients:** The autoregressive and moving average coefficients provide insights into the impact of past values and error terms on the current value, facilitating interpretation of the model's behavior.
* **Disadvantages of the ARIMA model:**
* **Complexity:** The ARIMA model can be more complex than individual AR or MA models, especially when higher-order models (p or q) and differencing (d) are required.
* **Parameter estimation:** Estimating the parameters (φ\_1, φ\_2, ..., φ\_p, θ\_1, θ\_2, ..., θ\_q) of the ARIMA model can be computationally intensive and may require iterative optimization algorithms.
* **Determining the model order:** Selecting the appropriate orders (p, d, q) of the ARIMA model can be challenging and often requires statistical techniques, model selection criteria, or domain knowledge.
* **Sensitivity to outliers:** The ARIMA model, similar to its AR and MA components, can be sensitive to outliers or extreme values in the data, potentially affecting its forecasting accuracy.
* **Applications of the ARIMA model:**
* **Time series forecasting:** ARIMA models are widely used for forecasting in various domains, including finance, economics, sales, and demand forecasting.
* **Anomaly detection:** ARIMA models can be utilized for detecting anomalies or unusual patterns in time series data, such as network traffic, system logs, or sensor readings.
* **Capacity planning:** ARIMA models find applications in capacity planning for infrastructure resources, such as server loads, network bandwidth, or power consumption.
* **Economic modeling:** ARIMA models are employed in economic analysis to forecast macroeconomic variables, analyze economic indicators, and support.
* **REFERENCES:** <https://towardsdatascience.com/time-series-models-d9266f8ac7b0>

**Q5. Write Short Note on:**

1. **White Noise:** Time series that show no autocorrelation is called white noise.

* A time series is white noise when sequence of uncorrelated random variables that are identically distributed White noise are variations in your data that cannot be explained by any regression model With a white noise, we cannot forecast future observations based on the past Whenever we do a time series analysis on a time series, we always assume that it is a combination is predictable and unpredictable components.
* **There are three ways to test if time series resembles white noise:** **1.** By plotting the time series, **2.** By comparing mean and standard deviation over time, **3.** By examining autocorrelation plots.
* Observing a time series data as a white noise can help to decide whether it can be used for or not Forecasting The concept of white noise is essential for time series analysis and forecasting, white noise tells you if you should further optimize the model or not.
* Time series data are expected to contain some white noise component on top of the signal generated by the underlying process.
* For example:

|  |  |
| --- | --- |
|  | **y(t) = signal(t) + noise(t)** |

* Once predictions have been made by a time series forecast model, they can be collected and analyzed. The series of forecast errors should ideally be white noise.
* When forecast errors are white noise, it means that all of the signal information in the time series has been harnessed by the model in order to make predictions. All that is left is the random fluctuations that cannot be modeled. A sign that model predictions are not white noise is an indication that further improvements to the forecast model may be possible.

1. **Stationarity:**

* A stationary time series's properties do not depend on the time at which the series is observed, for a wide-sense stationary time series, the mean and the variance/autocovariance keep constant over time.
* Differencing in statistics is a transformation applied to a non-stationary time-series in order to make it stationary in the mean sense (viz., to remove the non-constant trend) Differencing has nothing to do with the non-stationarity of the variance or autocovariance.

|  |
| --- |
| C:\Users\HP\Downloads\a1.jpg |

|  |
| --- |
| C:\Users\HP\Downloads\a2.jpg |

* Likewise, the seasonal differencing is applied to a seasonal time-series to remove the seasonal component. To difference the data, the difference between consecutive observations is computed. Mathematically, this is shown as
* Sometimes it may be necessary to difference the data a second time to obtain a stationary time series, which is referred to as second-order differencing:

1. **ACF:** The autocorrelation functions (ACF) is used for an estimate of q.

ACF stands for Autocorrelation Function, which is a statistical tool used in time series analysis to measure the correlation between observations of a time series at different lags. The ACF provides information about the linear relationship between a time series and its lagged values, which helps in understanding the underlying patterns and dependencies in the data.

* The Autocorrelation Function, denoted as ACF(k), is calculated for each lag k and measures the correlation between the time series observations at time t and the observations at time t-k. Mathematically, the ACF is defined as follows:

**ACF(k) = Cov(X\_t, X\_{t-k}) / Var(X\_t)**

* where X\_t represents the value of the time series at time t, Cov(X\_t, X\_{t-k}) is the covariance between X\_t and X\_{t-k}, and Var(X\_t) is the variance of X\_t.
* The ACF is typically visualized as a plot, called the autocorrelation plot or correlogram, where the lag (k) is plotted on the x-axis and the autocorrelation coefficient (ACF(k)) is plotted on the y-axis. The ACF plot provides insights into the correlation structure of the time series and helps in determining the presence of any significant lags or patterns.
* **Properties and Interpretation of ACF:**
* **Range and Interpretation:** The ACF values range from -1 to 1. A value of 1 indicates a perfect positive autocorrelation, -1 indicates a perfect negative autocorrelation, and 0 indicates no autocorrelation. Values between -1 and 1 represent varying degrees of correlation.
* **Significance:** In practice, the ACF plot often includes horizontal lines called confidence intervals. If an autocorrelation coefficient exceeds the confidence interval boundaries, it is considered statistically significant and indicates a correlation at that particular lag.
* **Decay:** The ACF plot shows the decay of autocorrelation as the lag increases. A rapid decay suggests a quick loss of correlation between observations, while a slow decay indicates persistence or long-term dependencies in the time series.
* **Seasonality:** In seasonal time series, the ACF plot often exhibits periodic patterns or spikes at multiples of the seasonal lag. This indicates the presence of seasonality in the data.
* **White Noise:** For a white noise time series, the ACF values are expected to be close to zero for all lags, indicating no significant correlation between observations at different time points.
* **Applications of ACF:**
* **Model Identification:** The ACF plot is commonly used in model identification and selection. It helps identify the order of autoregressive (AR) and moving average (MA) components in time series models, such as ARMA and ARIMA.
* **Seasonality Analysis:** The ACF plot helps identify the presence and lag of seasonality in time series data, aiding in the selection of appropriate seasonal models like SARIMA.
* **Residual Analysis:** ACF is used to analyze the autocorrelation structure of the residuals from a fitted model. It helps assess whether the model adequately captures the remaining correlations in the data.
* **Outlier Detection:** ACF can be used to detect outliers or influential observations in a time series by examining the autocorrelation values at different lags.

1. **PACF:**PACF stands for Partial Autocorrelation Function, which is a statistical tool used in time series analysis to measure the correlation between observations of a time series at different lags while controlling for the intermediate lags. The PACF provides insights into the direct relationship between a time series and its lagged values, helping to identify the order of the autoregressive (AR) component in a time series model.

* The Partial Autocorrelation Function, denoted as PACF(k), is calculated for each lag k and measures the correlation between the time series observations at time t and the observations at time t-k, while accounting for the influence of all intermediate lags. Mathematically, the PACF can be defined as follows:

**PACF(k) = Corr(X\_t, X\_{t-k} | X\_{t-1}, X\_{t-2}, ..., X\_{t-k+1})**

Where, X\_t represents the value of the time series at time t, X\_{t-k} represents the value at time t-k, and Corr() denotes the correlation.

* The PACF is typically visualized as a plot, known as the partial autocorrelation plot or PACF plot. In the PACF plot, the lag (k) is plotted on the x-axis, and the partial autocorrelation coefficient (PACF(k)) is plotted on the y-axis. The PACF plot helps in determining the appropriate lag order for the AR component of a time series model.
* **Properties and Interpretation of PACF:**
* **Initial Correlation:** The PACF at lag k is equal to the ACF at the same lag k, as both capture the direct correlation between the time series values at different lags.
* **Intermediate Lags:** The PACF eliminates the influence of the intermediate lags, providing a clearer picture of the direct relationship between the current observation and the observation at the given lag.
* **Significance:** Similar to ACF, the PACF plot often includes confidence intervals. PACF values exceeding the confidence interval boundaries are considered statistically significant, indicating a direct correlation at that specific lag.
* **AR Model Order**: The PACF plot helps identify the order of the AR component in a time series model. Significant PACF values at lag k indicate the need for including AR terms up to that lag in the model.
* **MA Component:** The PACF values beyond the significant lags typically diminish rapidly, suggesting the absence of a significant correlation beyond those lags. This behavior is characteristic of an AR(p) process without a significant moving average (MA) component.
* **Applications of PACF:**
* **Model Identification:** The PACF plot is crucial in model identification, particularly for determining the order of the AR component in ARMA and ARIMA models.
* **Model Selection:** The PACF helps in selecting the appropriate lag order for the AR component, enabling the construction of parsimonious and accurate time series models.
* **Residual Analysis:** The PACF is used to analyze the autocorrelation structure of the residuals from a fitted model, especially when assessing whether the model captures the remaining autocorrelation.
* **Time Series Forecasting:** Understanding the PACF aids in developing effective forecasting models by identifying the relevant AR terms to include in the model.

**(PPT Give you more understanding than PDF)** The material for the PDF has been compiled from various sources such as books, tutorials (offline and online), lecture notes, several resources available on Internet. The information contained in this PDF is for general information and education purpose only. While we endeavor to keep the information up to date and correct, **we make no representation of any kind about the completeness and accuracy of the material.** The information shared through this PDF material should be used for educational purpose only.

**Thank you…**