Time series analysis involves examining data points collected or recorded at specific time intervals. It is a powerful tool for understanding the underlying patterns, trends, and seasonal variations in the data, and for making forecasts about future values.

**Key Concepts in Time Series Analysis**

1. **Trend**: A long-term increase or decrease in the data.
2. **Seasonality**: Regular, repeating patterns or cycles in the data.
3. **Cyclical Patterns**: Fluctuations that are not of a fixed period, often influenced by economic or other factors.
4. **Noise**: Random variation in the data that cannot be attributed to trend, seasonality, or cycles.

**Steps in Time Series Analysis**

1. **Data Collection**: Gather data points recorded at consistent intervals (daily, monthly, yearly, etc.).
2. **Visualization**: Plot the time series to visually inspect patterns, trends, and anomalies.
3. **Decomposition**: Break down the time series into its components (trend, seasonality, and residual/noise).
4. **Stationarity Testing**: Use statistical tests to check if the series is stationary. Common tests include the Augmented Dickey-Fuller (ADF) test and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test.
5. **Differencing and Transformation**: Apply differencing or transformations to make the series stationary if necessary.
6. **Model Selection and Fitting**: Choose and fit an appropriate model to the data. Common models include ARIMA (AutoRegressive Integrated Moving Average), SARIMA (Seasonal ARIMA), and Exponential Smoothing.
7. **Model Evaluation**: Evaluate the model's performance using metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), or Root Mean Squared Error (RMSE).
8. **Forecasting**: Use the fitted model to make forecasts about future values.

**Common Time Series Models**

1. **ARIMA (AutoRegressive Integrated Moving Average)**:
   * Combines autoregressive (AR) terms, differencing (I), and moving average (MA) terms.
   * Suitable for non-seasonal data.
2. **SARIMA (Seasonal ARIMA)**:
   * Extends ARIMA by adding seasonal components.
   * Suitable for data with seasonality
3. **Exponential Smoothing**:
   * Models level, trend, and seasonality with different smoothing parameters.
   * Includes methods like Holt-Winters seasonal method.

**Practical Tips**

* **Data Preprocessing**: Ensure the data is clean and handle missing values appropriately.
* **Model Diagnostics**: Check residual plots to ensure no patterns are left in the residuals.
* **Multiple Models**: Compare different models to find the best fit for your data.
* **Domain Knowledge**: Incorporate knowledge about the data and domain to improve model accuracy.

**Stationary vs Non-Stationary Time Series:**

**Stationary Time Series**

A time series is considered **stationary** if its statistical properties do not change over time. This means the mean, variance, and autocorrelation structure remain constant throughout the series. Stationary time series are easier to model and predict as they follow a consistent pattern over time.

**Characteristics of a stationary time series:**

* **Constant mean**: The average value of the series remains the same over time.
* **Constant variance**: The spread of the series around the mean does not change over time.
* **Constant autocorrelation structure**: The relationship between current and past values of the series remains the same over time.

**Examples of stationary processes:**

* White noise series.
* A stationary ARMA (AutoRegressive Moving Average) process.

**Why is stationarity important?**

* Many time series models, like ARIMA (AutoRegressive Integrated Moving Average), require the series to be stationary.
* Stationarity simplifies the analysis and modeling process because the model parameters do not change over time.

**Non-Stationary Time Series**

A time series is considered **non-stationary** if its statistical properties change over time. This means the mean, variance, or autocorrelation structure varies over time, making it more challenging to model and predict.

**Characteristics of a non-stationary time series:**

* **Changing mean**: The average value of the series changes over time.
* **Changing variance**: The spread of the series around the mean changes over time.
* **Changing autocorrelation structure**: The relationship between current and past values of the series changes over time.

**Examples of non-stationary processes:**

* Trends: A consistent upward or downward movement in the series.
* Seasonal effects: Regular patterns that repeat over specific intervals.
* Random walk: A process where the value at time t is equal to the value at time t-1 plus a random shock.

**Handling non-stationary time series:**

* **Differencing**: Subtracting the previous observation from the current observation to remove trends.
* **Transformation**: Applying mathematical transformations (e.g., logarithm, square root) to stabilize variance.
* **Decomposition**: Separating the series into trend, seasonal, and residual components.

**Visual Examples**

1. **Stationary Time Series**:

*The mean and variance remain constant over time.*

1. **Non-Stationary Time Series (with trend)**:

*The mean increases over time, indicating a trend.*

1. **Non-Stationary Time Series (with seasonality)**:

*The series shows repeating patterns at regular intervals.*

Understanding whether a time series is stationary or non-stationary is a crucial first step in time series analysis, guiding the choice of appropriate modeling techniques.

**What is detrend a time series?**

Detrending a time series involves removing trends from the data to analyze the underlying patterns more clearly, such as seasonality or cycles. Trends in a time series can obscure these patterns and make it difficult to understand the true behavior of the data.

**Why Detrend a Time Series?**

1. **Focus on Other Components**: By removing the trend, you can focus on other components of the time series, such as seasonality or noise.
2. **Stationarity**: Many time series models assume that the data is stationary (i.e., its statistical properties do not change over time). Detrending helps achieve stationarity.
3. **Improved Forecasting**: Detrending can improve the accuracy of forecasting models by isolating patterns that are more consistent over time.

**Methods of Detrending**

There are several methods to detrend a time series, including:

1. **Differencing**:
   * **Method**: Subtract the previous observation from the current observation.
   * **Usage**: Often used to remove linear trends.
2. **Fitting a Regression Model**:

* **Method**: Fit a regression line (e.g., linear, polynomial) to the time series and then subtract the fitted values from the original data.
* **Usage**: Suitable for more complex trends.

1. **Moving Average**:

* **Method**: Apply a moving average to smooth the data and then subtract the smoothed values.
* **Usage**: Useful for removing short-term fluctuations

1. **Decomposition**:

* **Method**: Decompose the time series into trend, seasonal, and residual components, then remove the trend component.
* **Usage**: Provides a detailed breakdown of the time series.

**What is de-seasonalize a time series?**

De-seasonalizing a time series involves removing seasonal effects to better understand and analyze the underlying trends and other components of the data. Seasonal effects are patterns that repeat at regular intervals, such as daily, weekly, monthly, or yearly cycles. De-seasonalizing helps isolate the trend and cyclical components, making the time series more stationary and easier to model.

**Why Deseasonalize a Time Series?**

1. **Isolate Trends and Cycles**: Removing seasonal effects helps in identifying and analyzing long-term trends and cyclical behavior.
2. **Improve Forecasting Accuracy**: Many forecasting models assume the absence of seasonality. De-seasonalizing the series can improve the accuracy of these models.
3. **Stationarity**: De-seasonalizing helps achieve stationarity, a property required by many time series analysis methods.

**Methods of De-seasonalizing**

There are several methods to de-seasonalize a time series, including:

1. **Seasonal Decomposition**:
   * **Method**: Decompose the time series into trend, seasonal, and residual components using methods like classical decomposition or STL (Seasonal and Trend decomposition using Loess).
   * **Usage**: Subtract the seasonal component from the original series.
2. **Seasonal Indices**:

* **Method**: Calculate seasonal indices by averaging the values for each season, then use these indices to adjust the original series.
* **Usage**: Suitable when the seasonal pattern is consistent over time.

**Moving Averages**:

* **Method**: Apply a centered moving average to smooth out the seasonality, then divide the original series by this moving average.
* **Usage**: Useful for monthly or quarterly data with a consistent seasonal pattern.

Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) are tools used in time series analysis to understand the dependencies between observations at different lags.

**Autocorrelation Function (ACF)**

The ACF measures the correlation between observations of a time series separated by K time units, or lags. Essentially, it shows how the current value of the series is related to its past values. The ACF is useful for identifying the overall pattern and determining the nature of the time series.

**Properties of ACF:**

* Values range from -1 to 1.
* A value close to 1 indicates a strong positive correlation.
* A value close to -1 indicates a strong negative correlation.
* A value close to 0 indicates no correlation.

**Uses of ACF:**

* Identifying the presence of patterns such as seasonality.
* Determining the order of autoregressive models (AR) by examining the rate at which autocorrelations diminish.

**Partial Autocorrelation Function (PACF)**

The PACF measures the correlation between observations of a time series separated by K time units, but with the linear dependence of the observations at all shorter lags removed. In other words, PACF shows the direct relationship between an observation and its Kth lag, removing the effects of the intervening lags.

**Properties of PACF:**

* Similar to ACF, values range from -1 to 1.
* PACF can help identify the order of autoregressive models (AR) directly.

**Uses of PACF:**

* Determining the order of an autoregressive (AR) model by finding the lag beyond which PACF values become insignificantly different from zero.
* Helping to distinguish between AR and Moving Average (MA) processes.

**Key Differences:**

1. **ACF vs PACF for AR Processes:**
   * For an AR(p) process, the ACF shows a gradual decay, whereas the PACF cuts off after lag P.
2. **ACF vs PACF for MA Processes:**
   * For an MA(q) process, the ACF cuts off after lag q, whereas the PACF shows a gradual decay.
3. **ACF vs PACF for ARMA Processes:**
   * For an ARMA process, both the ACF and PACF exhibit a more complex pattern of decay.

**Practical Example:**

Imagine you have a time series of monthly sales data, and you want to build a model to predict future sales. By plotting the ACF and PACF of the series, you can:

* Use the ACF plot to identify seasonality and overall trends.
* Use the PACF plot to determine the appropriate number of lag terms to include in an AR model.

**Additive Time Series vs Multiplicative Time Series**:

In time series analysis, understanding the structure and components of a time series is crucial for accurate modeling and forecasting. Two common structures are additive and multiplicative time series. These structures describe how the components of a time series combine to form the observed data.

**Additive Time Series**

In an additive time series, the components (trend, seasonality, and noise) combine in a linear way. The basic form of an additive model is:

A white background with black text

Description automatically generated

**Characteristics of Additive Time Series:**

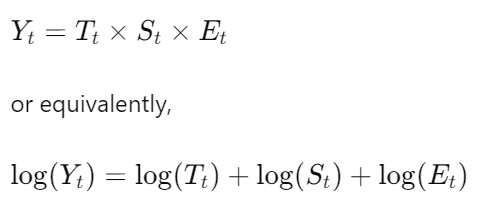
* The amplitude of the seasonal component remains constant over time.
* The model is suitable when the seasonal variations are roughly constant through the series.

**Example:**

* If you have monthly sales data with a consistent increase in sales over time and a stable seasonal pattern, an additive model may be appropriate.

**Multiplicative Time Series**

* In a multiplicative time series, the components combine in a multiplicative way. The basic form of a multiplicative model is:



**Characteristics of Multiplicative Time Series:**

* The amplitude of the seasonal component changes proportionally with the level of the series.
* The model is suitable when the seasonal variations are proportional to the level of the series.

**Example:**

If you have monthly sales data where the seasonal effect increases as the level of sales increases, a multiplicative model may be appropriate.

**Why make a non-stationary series stationary before forecasting?**

Making a non-stationary series stationary before forecasting is crucial in time series analysis for several reasons:

**1. Statistical Properties Consistency**

Stationarity means that the statistical properties of the series (mean, variance, autocorrelation, etc.) do not change over time. Most statistical models and forecasting techniques assume that the series is stationary. If these properties change, the model's predictions might become unreliable.

**2. Improved Model Accuracy**

Non-stationary data can lead to misleading correlations and trends. For example, a model might capture a trend as a significant relationship when it’s just a result of the non-stationarity. By making the series stationary, you remove these spurious relationships, allowing the model to capture the true underlying patterns.

**3. Simplification of the Modeling Process**

Stationary series are easier to model because their behavior is consistent over time. Many time series models, like ARIMA (AutoRegressive Integrated Moving Average), are designed to work with stationary data. These models use past values and their relationships to predict future values, which is only meaningful if these relationships are stable over time.

**4. Enhanced Predictive Performance**

Stationary time series models are more predictable and reliable. The model's parameters remain constant over time, leading to more accurate and consistent forecasts.

**Techniques to Make a Series Stationary**

1. **Differencing:** Subtracting the current observation from the previous one (or from a lagged value) can remove trends and make the series stationary.
2. **Transformation:** Applying mathematical transformations like logarithms or square roots can stabilize the variance.
3. **Decomposition:** Separating the time series into trend, seasonal, and residual components, and then modeling the stationary residual component.
4. **Detrending:** Removing a fitted trend line from the series.

**Key Points**

* **ADF Test:** The Augmented Dickey-Fuller (ADF) test is used to check if a time series is stationary. A low p-value (typically ≤ 0.05) indicates that the series is stationary.
* **Differencing:** This example shows how differencing can remove trends and make a series stationary. The differenced series is plotted to visualize the change.
* **Interpretation:** The ADF test results help confirm if the differencing made the series stationary.

By making a time series stationary, you ensure that your forecasting model is based on stable and reliable relationships, leading to more accurate and meaningful predictions.

**Questions & Answers:**

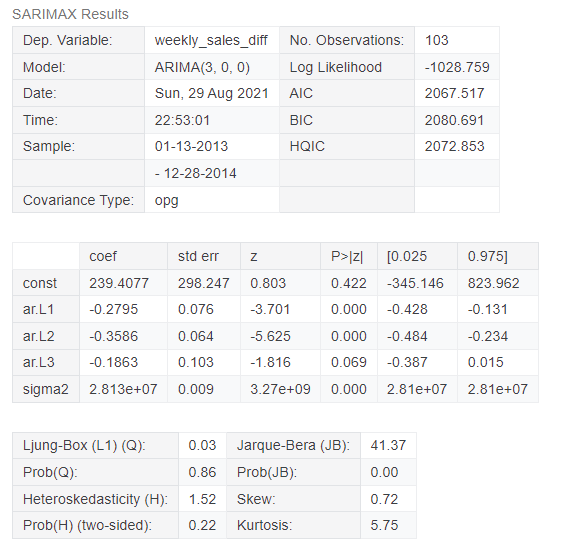
* 1. **In Naive's Forecast, Simple moving Average, Simple Exponential Smoothing, Double Exponential Smoothing, Triple Exponential Smoothing. Do we need to check the stationarity of the time series data?**

**Summary 🡪**

* **Naive Forecast:** No need to check for stationarity.
* **Simple Moving Average:** No strict need to check for stationarity, but it works better with stationary data.
* **Simple Exponential Smoothing:** Assumes data is stationary (no trend or seasonality).
* **Double Exponential Smoothing:** No need to check for stationarity; handles trends.
* **Triple Exponential Smoothing:** No need to check for stationarity; handles trends and seasonality.

In general, while stationarity is a crucial consideration for certain time series models like ARIMA, it is less critical for the smoothing methods mentioned here, especially those designed to handle trends and seasonality explicitly. However, understanding the characteristics of your data and pre-processing it appropriately can always help improve the performance of your forecasting models.

* 1. **What below indicates**



**Dep. Variable: weekly\_sales\_diff :** The dependent variable being modeled is weekly\_sales\_diff, which likely represents the differenced weekly sales data to achieve stationarity.

**No. Observations: 103: 🡪** The number of observations in the dataset is 103.

**Model: ARIMA(3, 0, 0) 🡪** This specifies that an ARIMA model with 3 autoregressive terms, 0 differences, and 0 moving average terms was used.

**Log Likelihood: -1019.340 🡪** The log likelihood of the fitted model. Higher (less negative) values indicate a better fit to the data.

**AIC: 2048.679 🡪** Akaike Information Criterion. A lower AIC indicates a better fit, balancing model complexity and goodness of fit.

**BIC: 2061.853 🡪** Bayesian Information Criterion. Like AIC but includes a penalty for the number of parameters in the model. Lower BIC is better.

**HQIC: 2054.015 🡪** Hannan-Quinn Information Criterion. Like AIC and BIC but with a different penalty term.

**Covariance Type: opg 🡪** The type of covariance matrix used for standard error estimation, in this case, the outer product of gradients.

**Interpretation of Coefficients:**

* **const (constant term)**: The constant term is approximately 27,210, indicating the mean level of the differenced series.
* **ar.L1**: The coefficient for the first lag of the autoregressive term is 0.5407, suggesting a positive relationship with the first lag.
* **ar.L2**: The coefficient for the second lag is -0.1536, suggesting a slight negative relationship, but the p-value (0.137) indicates it is not statistically significant at the 0.05 level.
* **ar.L3**: The coefficient for the third lag is 0.1456, suggesting a positive relationship, but again the p-value (0.164) indicates it is not statistically significant at the 0.05 level.
* **sigma2 (variance of the residuals)**: The variance of the residuals is very large, indicating high variability in the data.

**Statistical Tests and Diagnostics**

* **Ljung-Box (L1) (Q): 0.01 🡪** A test for autocorrelation at lag 1. The very low Q value suggests no significant autocorrelation at this lag.
* **Prob(Q): 0.90 🡪** The p-value for the Ljung-Box test. A high p-value (0.90) suggests the null hypothesis of no autocorrelation cannot be rejected.
* **Jarque-Bera (JB): 193.47 🡪** A test for normality of residuals. A high JB value indicates non-normality.
* **Prob (JB): 0.00 🡪** The p-value for the Jarque-Bera test. A p-value of 0.00 indicates strong evidence against the null hypothesis of normality.
* **Heteroskedasticity (H): 1.39 🡪** A test for constant variance (homoscedasticity) in the residuals. A value close to 1 suggests homoscedasticity.
* **Prob(H) (two-sided): 0.34 🡪** The p-value for the heteroskedasticity test. A p-value of 0.34 indicates no significant evidence of heteroskedasticity.
* **Skew: 1.50 🡪** The skewness of the residuals. A value of 1.50 indicates a positive skew.
* **Kurtosis: 9.01 🡪** The kurtosis of the residuals. A value of 9.01 indicates heavy tails (leptokurtic distribution).
  1. **ARIMA vs SARIMA?**

**ARIMA Model:**

* ARIMA stands for AutoRegressive Integrated Moving Average.
* It models non-seasonal time series data.
* The model is specified with three parameters: (p, d, q), where:
  + p is the order of the autoregressive part.
  + d is the order of differencing.
  + q is the order of the moving average part.

**Seasonal ARIMA (SARIMA) Model:**

* SARIMA stands for Seasonal AutoRegressive Integrated Moving Average.
* It extends ARIMA by explicitly including seasonal components.
* The model includes both non-seasonal and seasonal parameters: (p, d, q) and (P, D, Q, s), where:
  + P is the order of the seasonal autoregressive part.
  + D is the order of seasonal differencing.
  + Q is the order of the seasonal moving average part.
  + s is the length of the seasonal cycle (e.g., 12 for monthly data with yearly seasonality).
  1. **Why Seasonality is considered in SARIMA:**

1. **Capturing Different Patterns:**

* ARIMA models are suitable for capturing short-term patterns and trends in time series data.
* However, when there are clear seasonal patterns (e.g., monthly sales data showing yearly seasonality), ARIMA alone might not be sufficient.
* SARIMA incorporates both short-term (non-seasonal) and long-term (seasonal) components, allowing it to capture complex patterns in the data.

1. **Seasonal Differencing:**

* Seasonal differencing (represented by D) is used to remove seasonality in the data, making it stationary. This is particularly useful for time series with strong seasonal components.
* For instance, if the data shows a yearly seasonality, seasonal differencing will subtract the value from the same period in the previous year, helping to stabilize the mean of the series.

1. **Seasonal AR and MA Components:**

* Seasonal autoregressive (AR) and moving average (MA) components (represented by P and Q) capture the dependencies at seasonal lags.
* For example, in monthly data with yearly seasonality, the seasonal AR component would account for the relationship between observations 12 months apart.

**Difference between Augmented Engle-Granger vs** **Augmented Dickey-Fuller (ADF):**

|  |  |  |
| --- | --- | --- |
|  | **Augmented Dickey-Fuller (ADF)** | **Augmented Engle-Granger** |
| **Objective** | Tests for the presence of a unit root (non-stationarity) in a single time series. | Tests for cointegration between two or more time series by checking the stationarity of the residuals from a cointegrating regression. |
| **Context** | Used when dealing with a single time series to determine if it is stationary or needs differencing. | Used when dealing with multiple time series to determine if they share a common stochastic trend (i.e., are cointegrated). |
| **Procedure** | Directly applied to a single time series | Involves two steps - first, perform a regression to obtain residuals; second, apply the ADF test to these residuals. |
| **Null Hypothesis** | * 1. If the test statistic is less than the critical value, reject the null hypothesis (the series is stationary).   2. If the test statistic is greater than the critical value, fail to reject the null hypothesis (the series is non-stationary). | * 1. If the p-value of the ADF test on the residuals is greater than the critical value, null hypothesis (no cointegration between series) must be considered.   2. If the p-value is less than the critical value, reject the null hypothesis (cointegration exists). |

**How Facebook Prophet model identifies the seasonality and trend in time series data?**