8.03 ARIMA

Recap: Time Series

- A time series is a measure of unit change over time for any variables under observation

- The built-in Python datetime module supplies classes for manipulating dates and times in both simple and complex ways

Time Series Analysis

- Time series analysis comprises methods for analyzing temporal data in order to extract meaningful statistics and other characteristics of the data.

- Time series forecasting is the use of a model to predict future values based on previously observed values

- Time series are widely used for non-stationary data, like economic, weather, stock price, and retail sales in this post.

Time Series Components

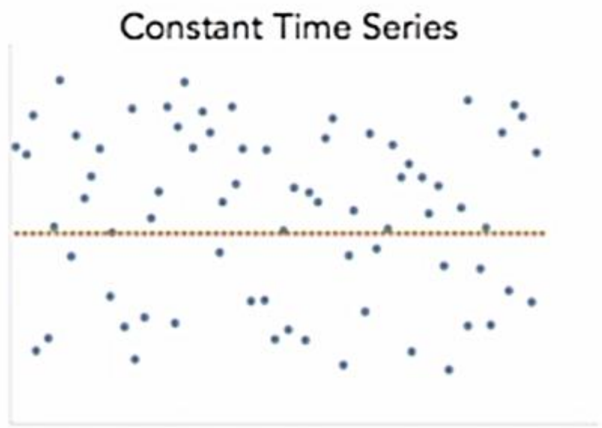
1. **Trend:** Upward, downward, or stationary. If your company sales increase every year, it is showing an upward trend.

2. **Seasonality:** Repeating pattern in certain period. Ex: difference between summer and winter. Also includes special holidays

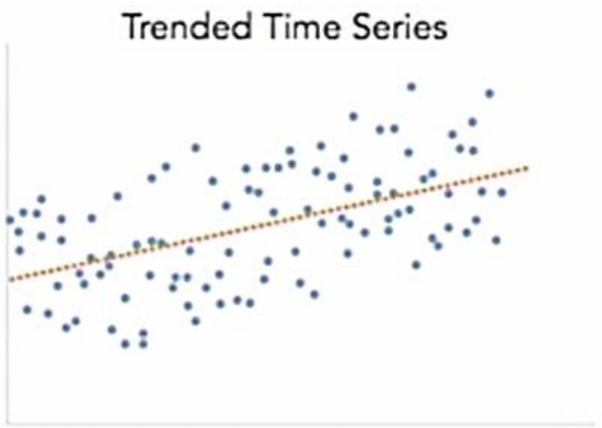
3. **Irregularity:** External factors that affect time series data such as Covid, natural disasters.

4. **Cyclic:** repeating up and down time series data.

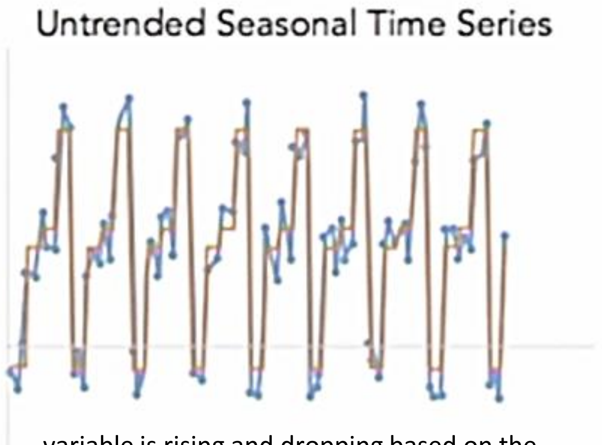
Time Series: Trends & Seasonality



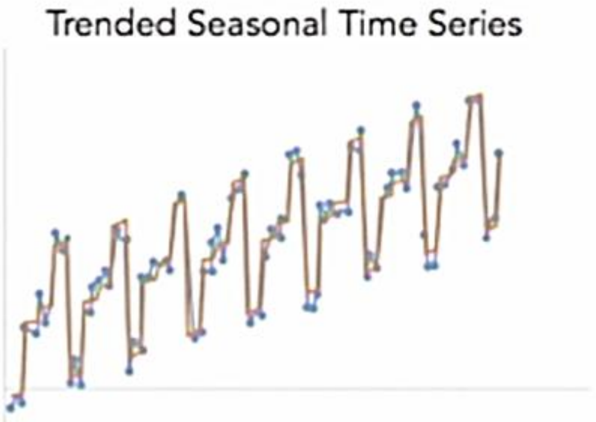
No trends or changes in variable over time



Net increase (or decrease) in variable over time



variable is rising and dropping based on the seasons in a year **WITHOUT** a net change in the average value of the time series



variable is rising and dropping based on the seasons in a year **WITH** a net change in the average value of the time series

ARIMA Overview

- ARIMA (autoregressive integrated moving average) is a regression model that uses time series data to either better understand the data set or to predict future trends
- A statistical model is autoregressive if it predicts future values based on past values
- For example, an ARIMA model might seek to predict a stock's future prices based on its past performance or forecast a company's earnings based on past periods

ARIMA Assumptions

- ARIMA a naive model, it assumes time series data we are working with satisfies following conditions:

  1. "non-seasonal" meaning different seasons do not affect its values. When there exists seasonality, we use SARIMA short for Seasonal ARIMA model

  2. No Irregularity. Ex: No irregular events like COVID-19 that affect our data

ARIMA Components

- **Autoregressive (AR):** refers to a model that shows a changing variable that regresses on its own lagged, or prior, values.

- **Integrated (I):** represents the differencing of raw observations to allow for the time series to become stationary (i.e., data values are replaced by the difference between the data values and the previous values).

- **Moving average (MA):** incorporates the dependency between an observation and a residual error from a moving average model applied to lagged observations

ARIMA Components – **Autoregressive (AR):**

- Auto Regressive (AR) property of ARIMA is referred to as **P**.
- Past time points of time series data can impact current and future time points. ARIMA models take this concept into account when forecasting current and future values. ARIMA uses several lagged observations of time series to forecast observations. A weight is applied to each of the past term and the weights can vary based on how recent they are.
- **AR(x)** means **x** lagged error terms are going to be used in the ARIMA model.
- ARIMA relies on AutoRegression. Autoregression is a process of regressing a variable on past values of itself. Autocorrelations gradually decay and estimate the degree to which white noise characterizes a series of data.

ARIMA Components – **Integrated (I):**

- If a trend exists, then time series is considered non-stationary and shows seasonality.
- Integrated is a property that reduces seasonality from a time series.
- ARIMA models have a degree of differencing which eliminates seasonality.
- I refers to D property in Python

ARIMA Components – **Moving average (MA):**

- Error terms of previous time points are used to predict current and future point’s observation. Moving average (MA) removes non-determinism or random movements from a time series. The property **Q** represents Moving Average in ARIMA. It is expressed as MA(x) where x represents previous observations that are used to calculate current observation.

- Moving average models have a fixed window and weights are relative to the time. This implies that the MA models are more responsive to current event and are more volatile.

ARIMA Parameters

- Each component in ARIMA functions as a parameter with a standard notation.
- For ARIMA models, a standard notation would be ARIMA with p, d, and q, where integer values substitute for the parameters to indicate the type of ARIMA model used.
- In a linear regression model, for example, the number and type of terms are included. A 0 value, which can be used as a parameter, would mean that particular component should not be used in the model. This way, the ARIMA model can be constructed to perform the function of an ARMA model, or even simple AR, I, or MA models.

ARIMA Parameters

The parameters can be defined as:

1. p: the number of lag observations in the model; also known as the lag order.

2. d: the number of times that the raw observations are differenced; also known as the degree of differencing.

3. q: the size of the moving average window; also known as the order of the moving average.

ARIMA Parameters – p

- Number of lags of Y to be used as predictors.

- In other words, If you are trying to predict June’s sale how many previous(lag) month’s data are you going to use?

ARIMA Parameters – d

- Minimum number of differencing needed to make time series data **stationary**.

- Already stationary data would have **d** = 0.

ARIMA Parameters – q

- The size of the moving average window; also known as the order of the moving average

ARIMA & Stationarity

In an autoregressive integrated moving average model, the data are differenced in order to make it stationary. A model that shows stationarity is one that shows there is constancy to the data over time. Most forms of data show trends, so the purpose of differencing is to remove any trends or seasonal structures.

Seasonality, or when data show regular and predictable patterns that repeat over a calendar year, could negatively affect the regression model. If a trend appears and stationarity is not evident, many of the computations throughout the process cannot be made with great efficacy.

ARIMA FAQs

What is ARIMA used for?

- ARIMA is a method for forecasting or predicting future outcomes based on a historical time series. It is based on the statistical concept of serial correlation, where past data points influence future data points.

ARIMA FAQs

What are the differences between autoregressive and moving average models?

- ARIMA combines autoregressive features with those of moving averages. An AR(2) autoregressive process, for instance, is one in which the current value is based on the previous two values.
- A moving average is a calculation used to analyze data points by creating a series of averages of different subsets of the full data set in order to smooth out the influence of outliers.
- As a result of this combination of techniques, ARIMA models can take into account trends, cycles, seasonality, and other non-static types of data when making forecasts.

ARIMA FAQs

How does ARIMA forecasting work?

- ARIMA forecasting is achieved by plugging in time series data for the variable of interest.
- Statistical methods will identify the appropriate number of lags or amount of differencing to be applied to the data and check for stationarity.
- It will then output the results, which are often interpreted similarly to that of a multiple linear regression model.

ARIMA FAQs

What does Stationary mean?

- Time series data considered stationary if it contains:
  - constant mean
  - constant variance
  - Covariance that is independent of time

ARIMA Applications **& Limitations**

- The model's goal is to predict outcomes by examining the differences between historical values in the series instead of via actual values
- They are widely used in technical analysis to forecast stock prices
- ARIMA requires stationary data and assumes future will resemble the past.
    - Hence, they can prove inaccurate under certain market conditions, such as financial crises or periods of rapid technological change.

Seasonal ARIMA (SARIMA)

- When dealing with seasonal effects, we make use of the seasonal ARIMA, which is denoted as ARIMA(p,d,q)(P,D,Q)s

- Here, (p, d, q) are the non-seasonal parameters described above, while (P, D, Q) follow the same definition but are applied to the seasonal component of the time series

- The term ‘s’ is the periodicity of the time series (4 for quarterly periods, 12 for yearly periods, etc.)

Seasonal ARIMA (SARIMA)

- The seasonal ARIMA method can appear daunting because of the multiple tuning parameters involved.

- When we review today's lesson code, we will review how to automate the process of identifying the optimal set of parameters for SARIMA

Seasonal ARIMA (SARIMA)

- Model notation is SARIMA(p, d, q).(P,D,Q)m
- These parameters account for seasonality, trend, and noise in data

AIC (Akaike information criterion)

- AIC is an estimator of the relative quality of ARIMA models for a given set of data
- Given a collection of models for the data, AIC estimates the quality of each model, relative to each of the other models.
- The lower the AIC value, the better

AIC (Akaike information criterion)

- The AIC measures how well a model fits the data while taking into account the overall complexity of the model.

- A model that fits the data very well while using lots of features will be assigned a larger AIC score than a model that uses fewer features to achieve the same goodness-of-fit.

- Therefore, we are interested in finding the model that yields the lowest AIC value.