Data Science

Module-11

ARIMA Model

Agenda

01 WHY ARIMA MODEL?

02 WHAT IS ARIMA MODEL?

03 ASSUMPTIONS OF ARIMA MODEL

04 STEPS TO BUILD ARIMA MODEL

Why ARIMA Model?

Why ARIMA Model

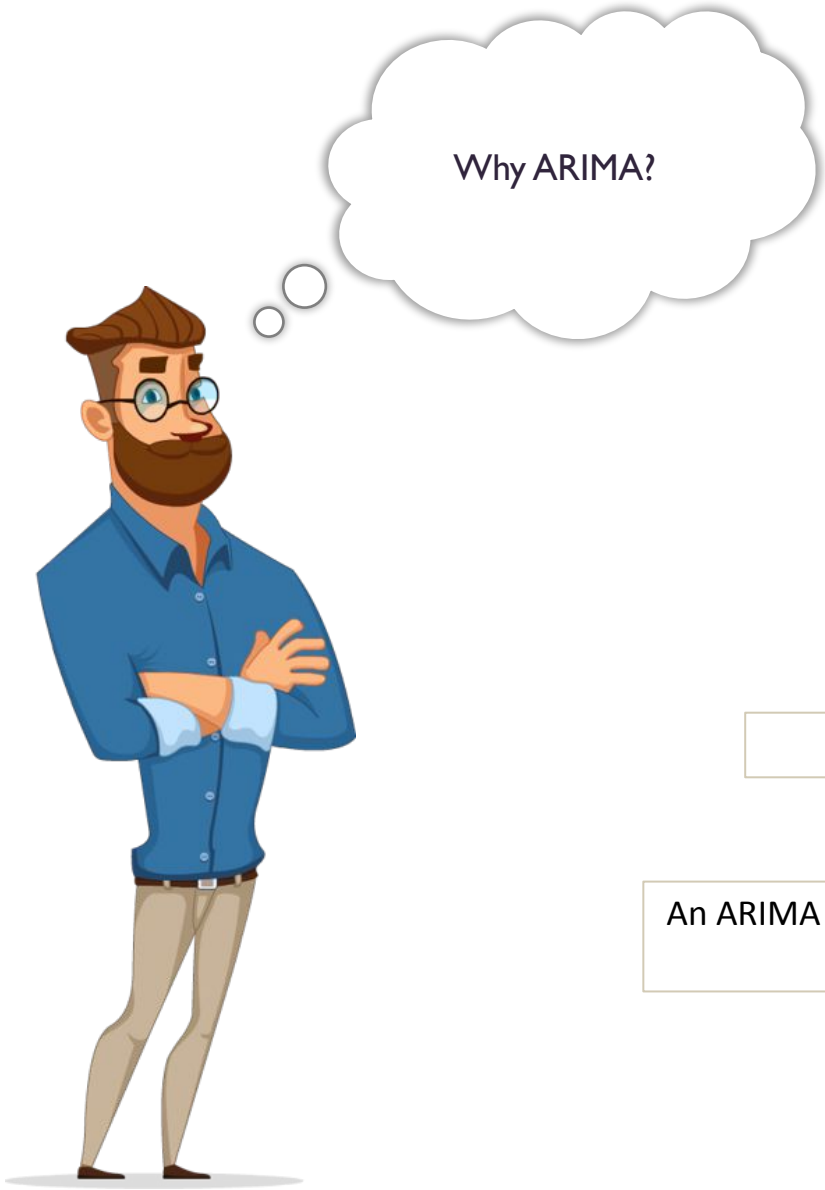
A class of statistical model for analyzing and forecasting time series data

ARIMA - **A**uto**R**egressive **I**ntegrated **M**oving **A**verage

Data show evidence of non-stationarity

A random variable that is a time series is stationary if its statistical properties are all constant over time.

Why ARIMA Model



A stationary series has **no trend**, its variations around its mean have a **constant amplitude**, and it wiggles in a consistent fashion

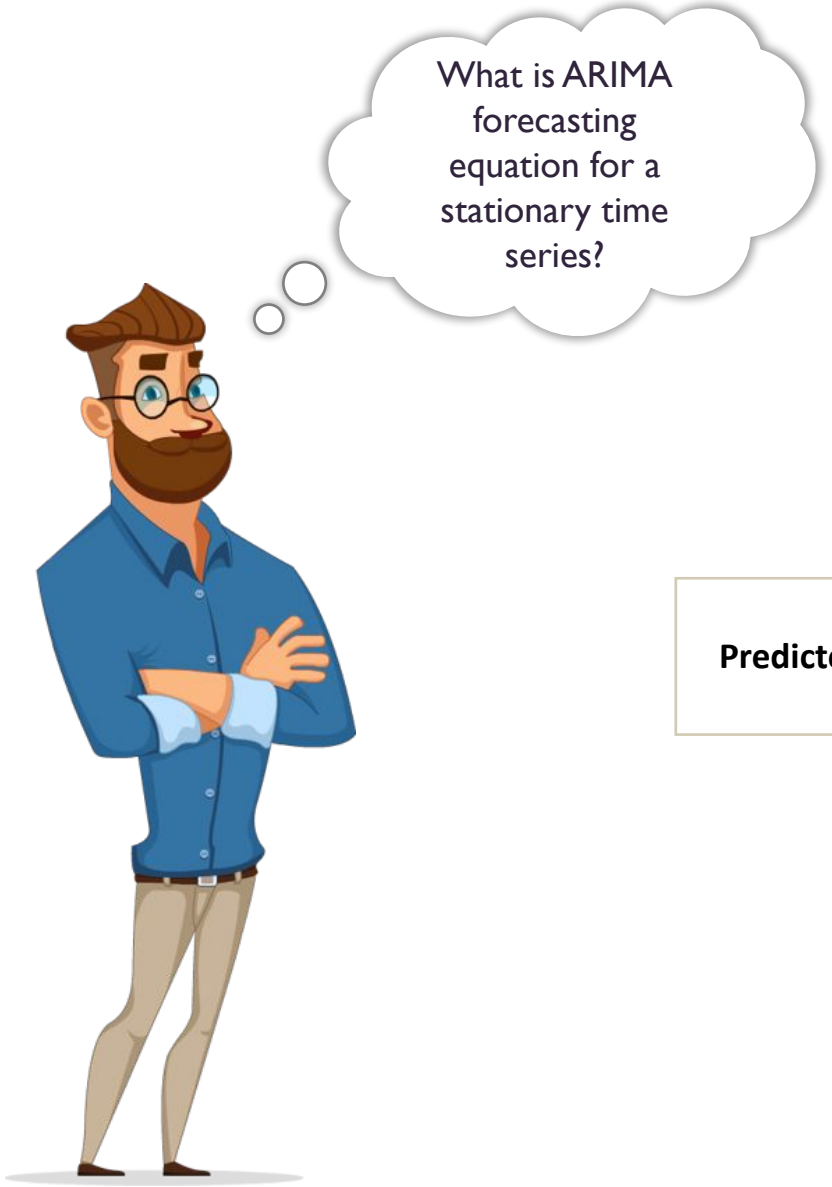
The latter condition means that its **autocorrelations** remain constant over time

A random variable of this form can be viewed as a combination of **signal and noise**

An ARIMA model can be viewed as a “filter” that tries to separate the signal from the noise, and the signal is then extrapolated into the future to obtain forecasts.

What is ARIMA Model?

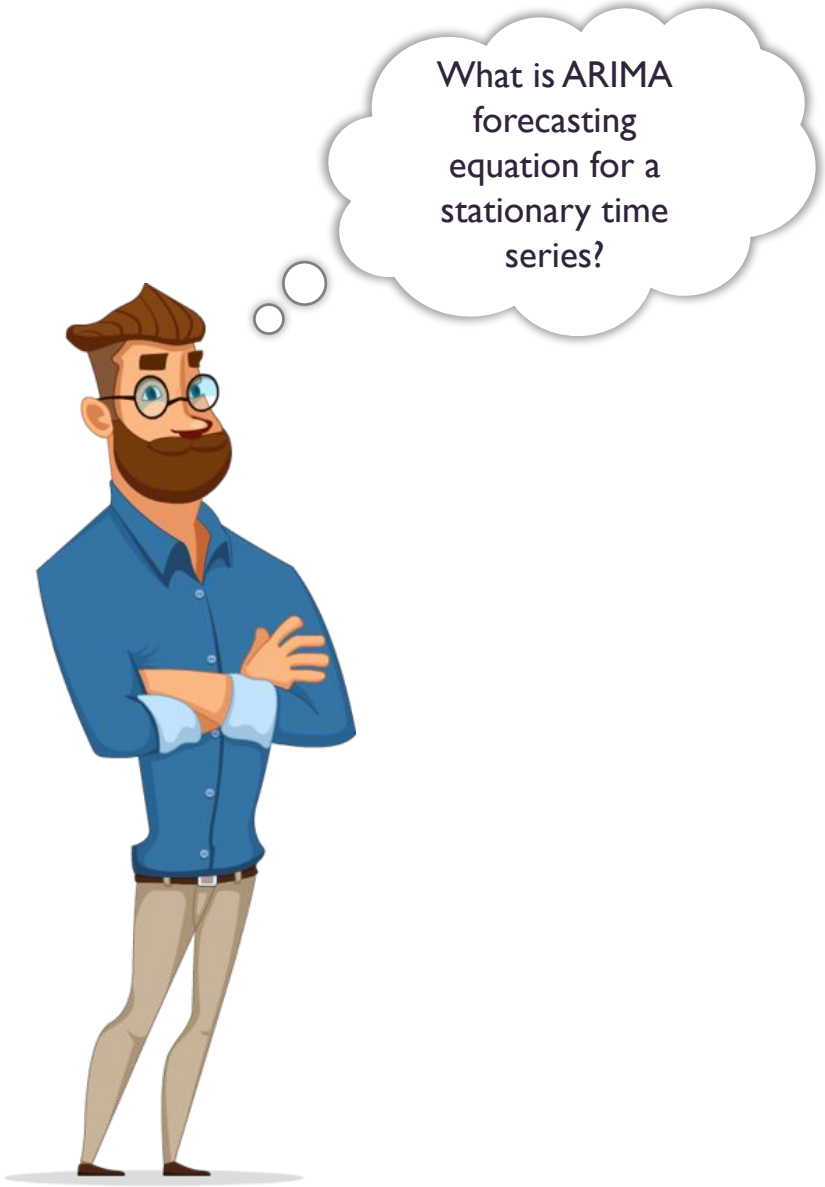
What is ARIMA Model?



A linear equation in which the predictors consist of lags of the dependent variable and/or lags of the forecast errors

Predicted value of Y =
- **a constant**
- **a weighted sum of one or more recent values of Y**
- **a weighted sum of one or more recent values of the errors.**

What is ARIMA Model?



AR = Auto**R**egressive - uses the dependent relationship between an observation and some number of lagged observations.

p = Lag order.

I = Integrated - The use of differencing of raw observations

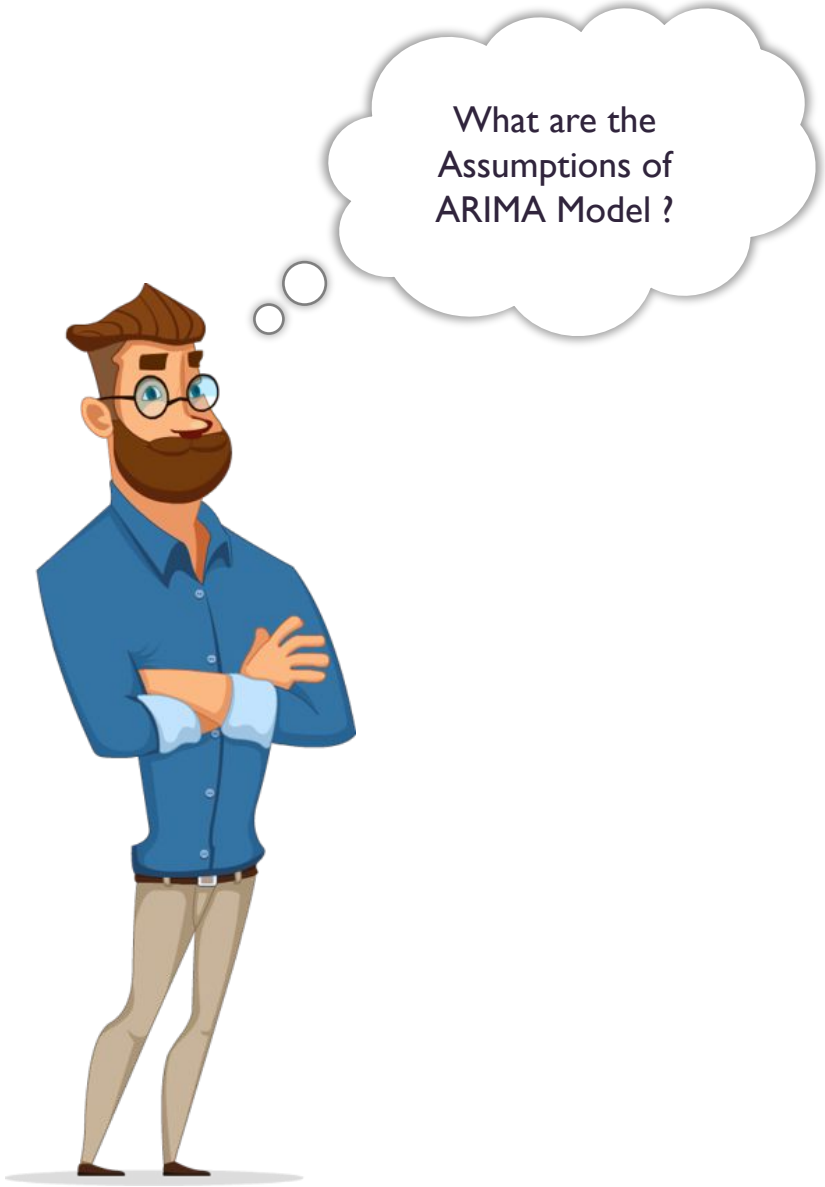
d = Degree of differencing

MA = Moving **A**verage - uses the dependency between an observation and residual errors from a moving average model applied to lagged observations.

q = Order of moving average

Assumptions of ARIMA Model

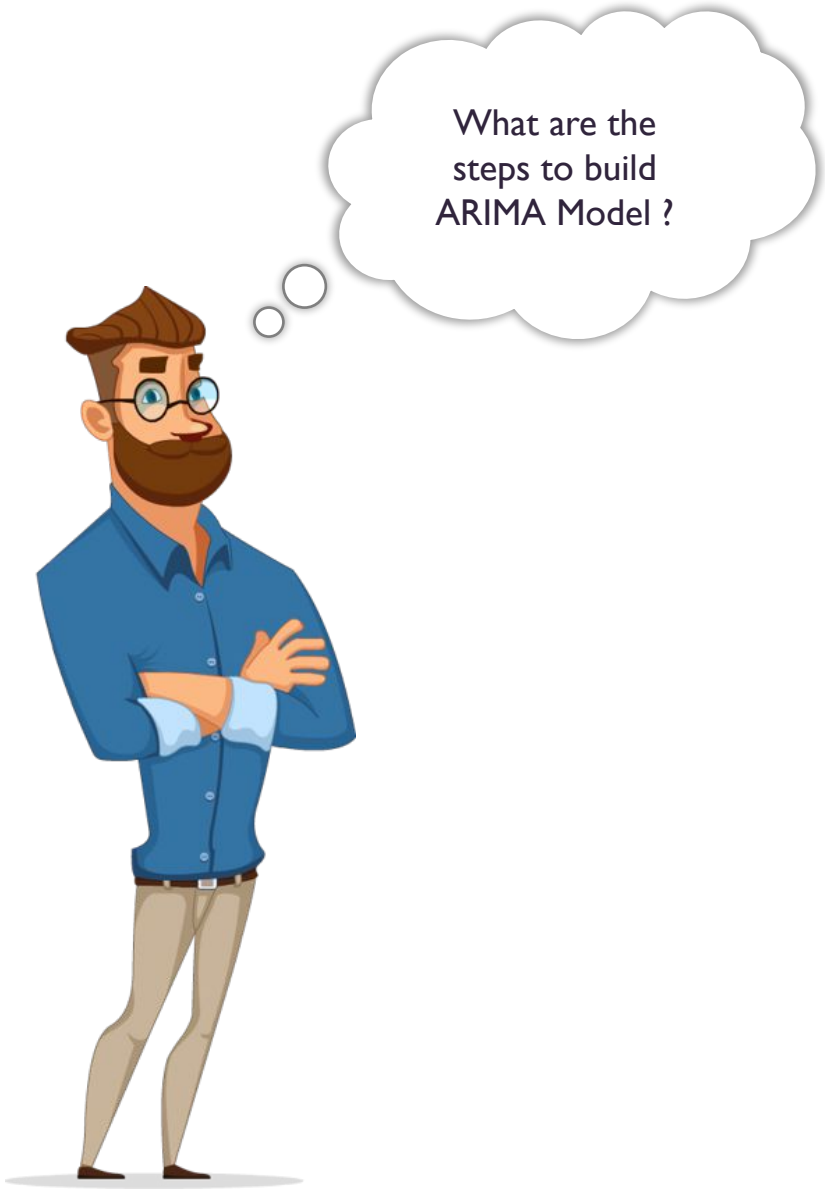
Assumptions of ARIMA Model



- **Stationarity**
- **Uncorrelated random error**
- **No outliers**
- **Random shocks (a random error component)**

Steps to build ARIMA Model

Steps to build ARIMA Model

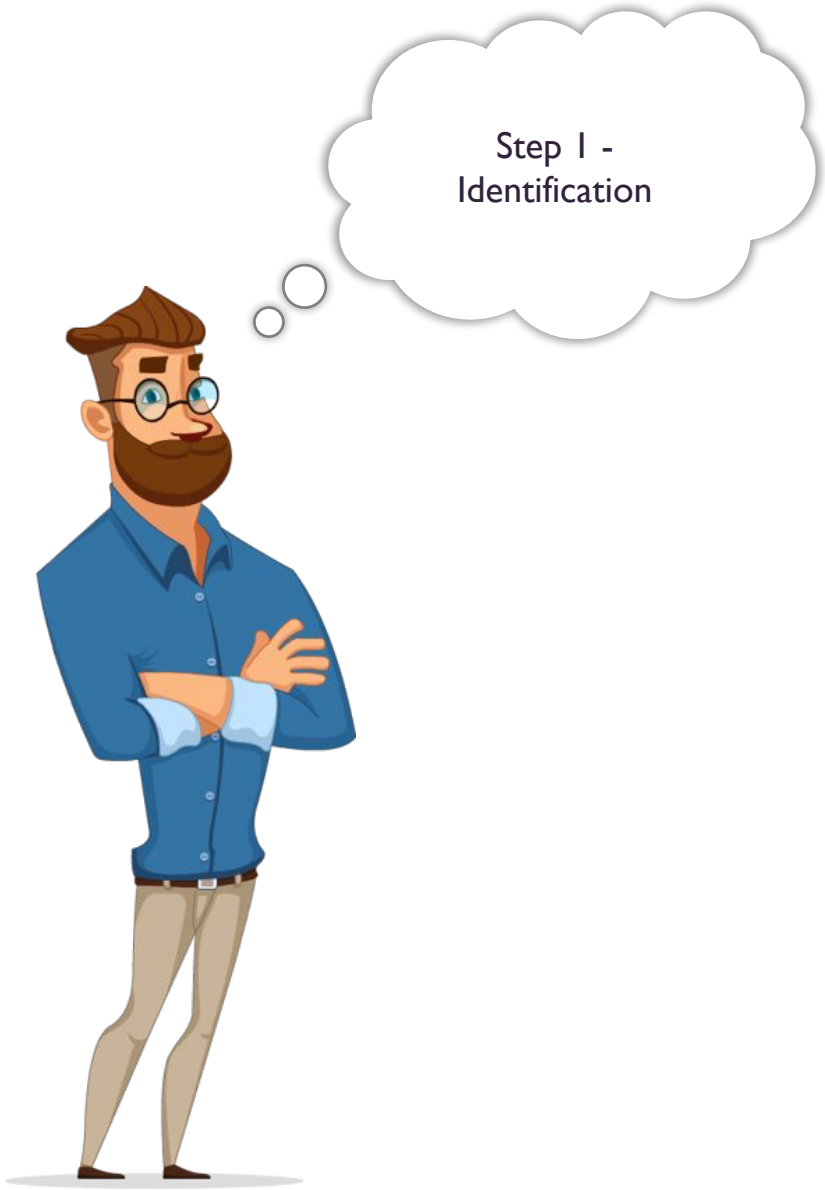


Box-Jenkins Method

An iterative approach consists of 3 steps

Identification → Estimation → Diagnostic Checking

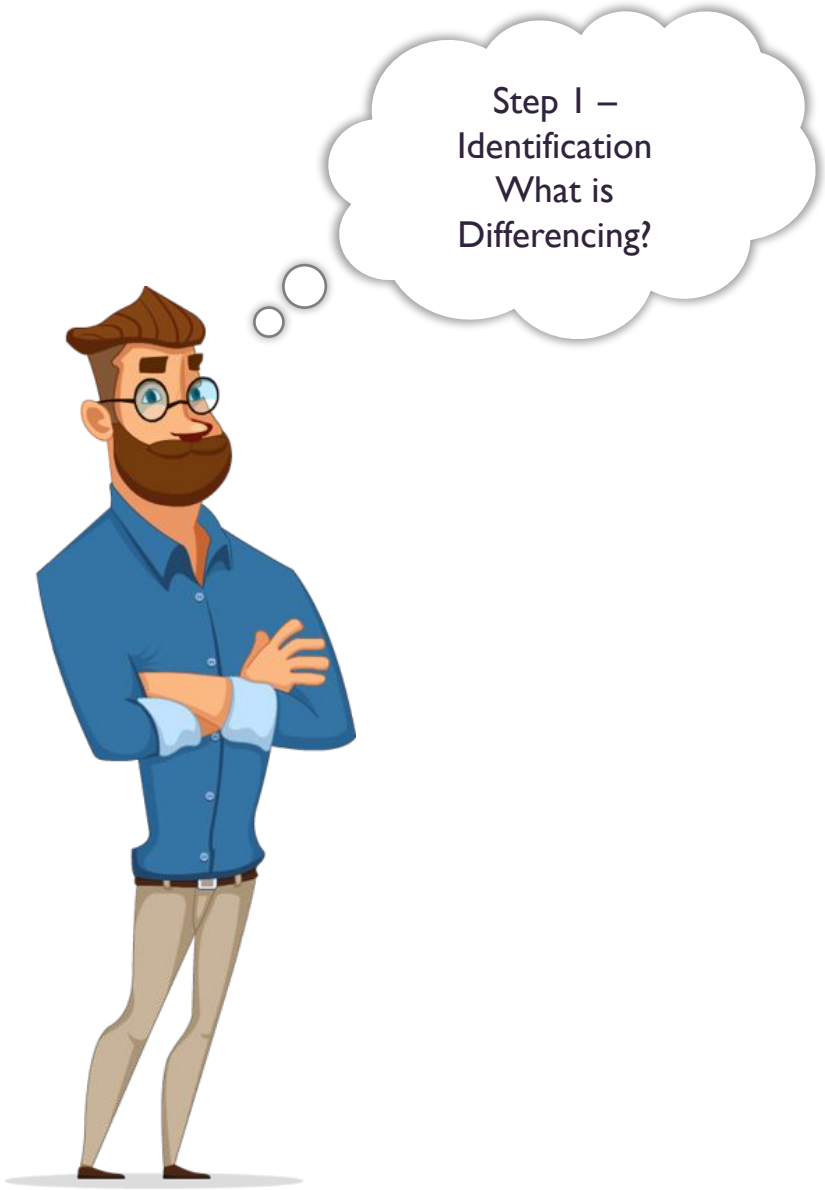
Steps to build ARIMA Model



Assess whether the time series is stationary, and if not, how many differences are required to make it stationary

Identify the parameters of an ARMA model for the data

Steps to build ARIMA Model



Unit Root Tests - to determine whether or not it is stationary

Avoid over differencing

Steps to build ARIMA Model

Autocorrelation Function (ACF).

The plot summarizes the correlation of an observation with lag values. The x-axis shows the lag and the y-axis shows the correlation coefficient between -1 and 1 for negative and positive correlation.

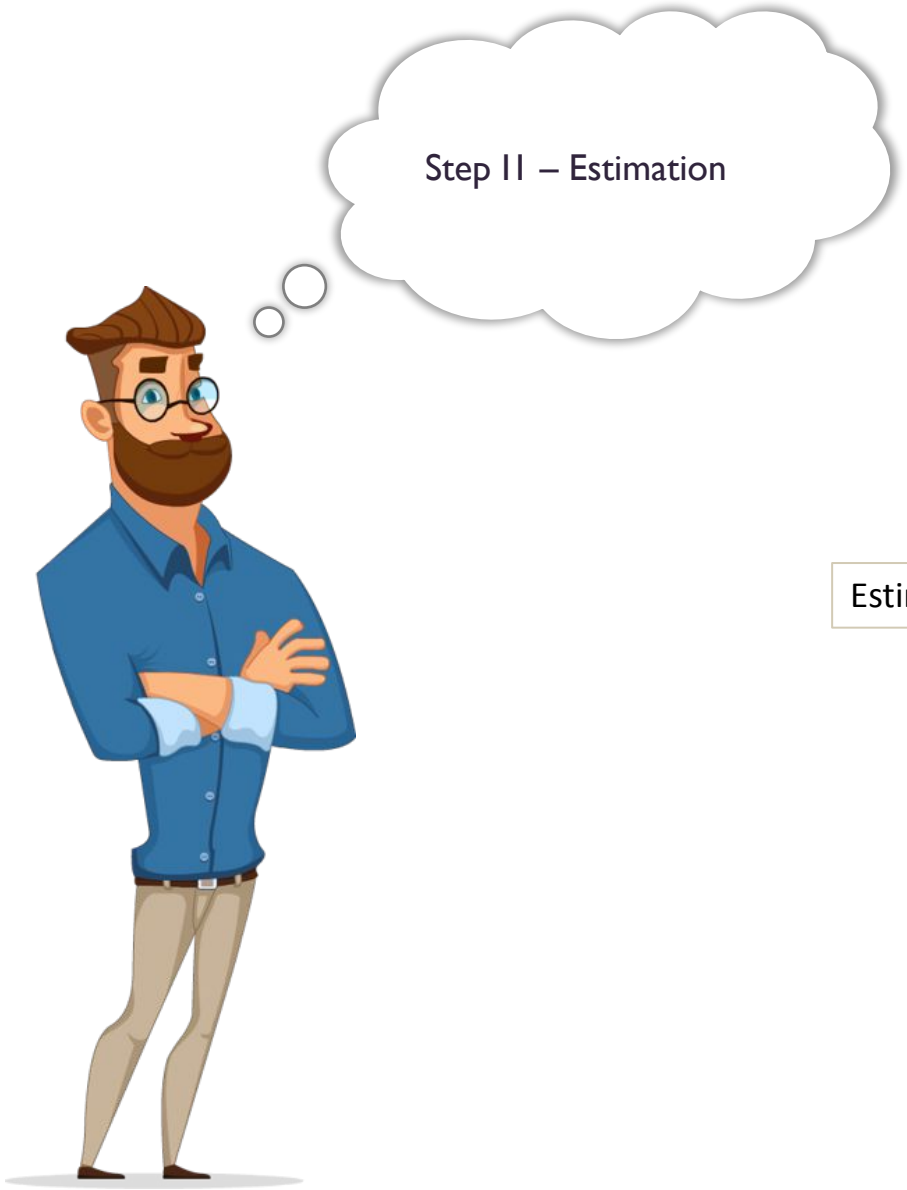
Partial Autocorrelation Function (PACF)

The plot summarizes the correlations for an observation with lag values that is not accounted for by prior lagged observations.

Steps to build ARIMA Model

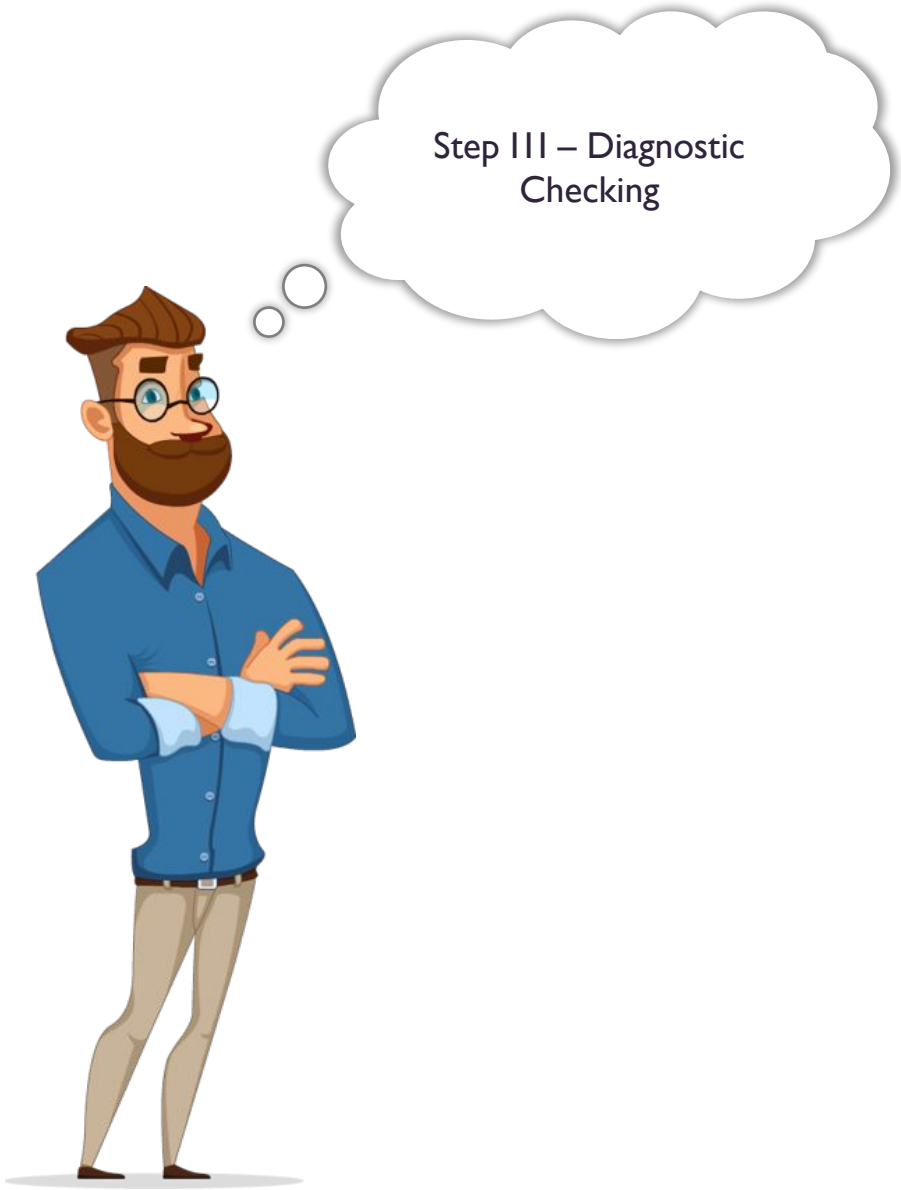
- The model is AR if the ACF trails off after a lag and has a hard cut-off in the PACF after a lag. This lag is taken as the value for p.
- The model is MA if the PACF trails off after a lag and has a hard cut-off in the ACF after the lag. This lag value is taken as the value for q.
- The model is a mix of AR and MA if both the ACF and PACF trail off.

Steps to build ARIMA Model



Estimation involves using numerical methods to minimize a loss or error term.

Steps to build ARIMA Model

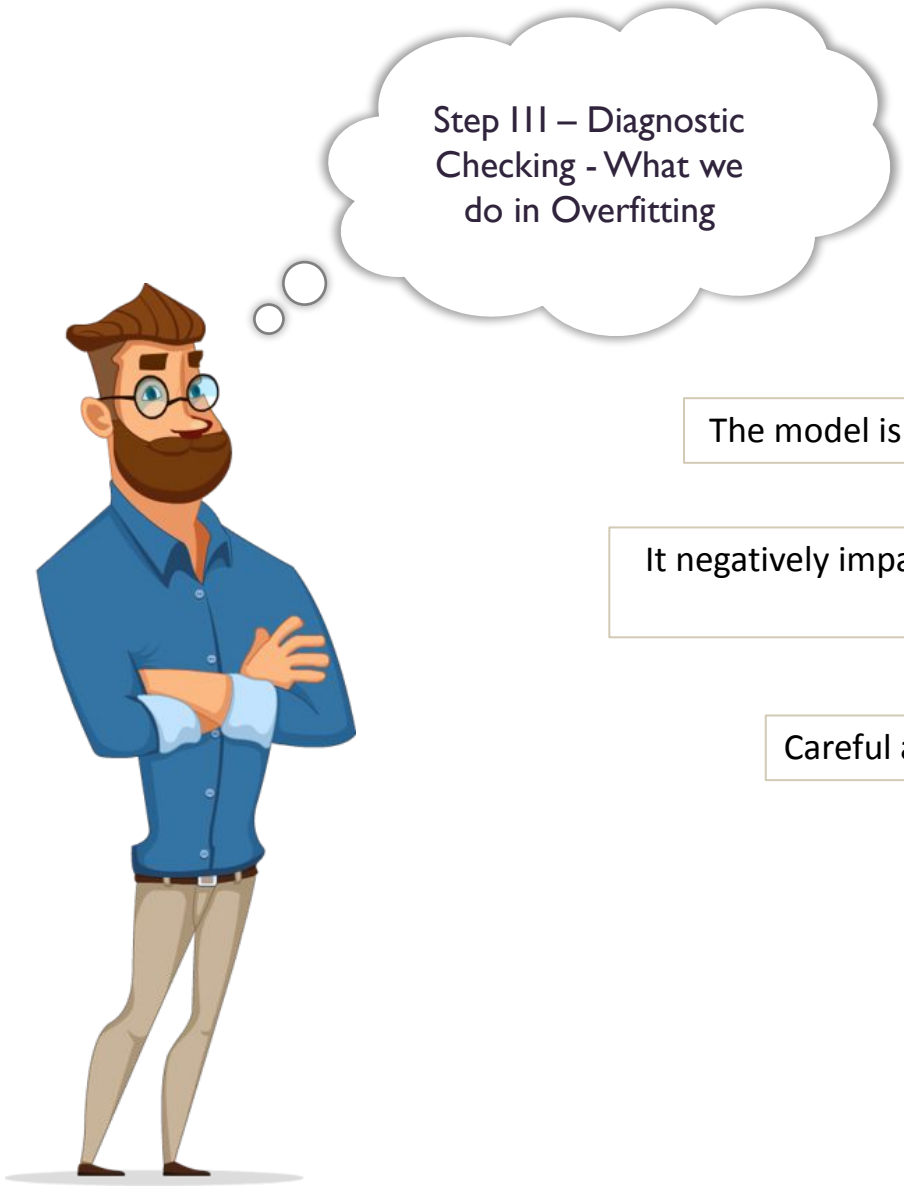


Look for evidence that the model is not a good fit for the data

Overfitting

Residual Errors

Steps to build ARIMA Model



The model is more complex than it needs to be and captures random noise in the training data.

It negatively impacts the ability of the model to generalize, resulting in poor forecast performance on out of sample data.

Careful attention must be paid to both in-sample and out-of-sample performance

Steps to build ARIMA Model

Forecast residuals provide a great opportunity for diagnostics.

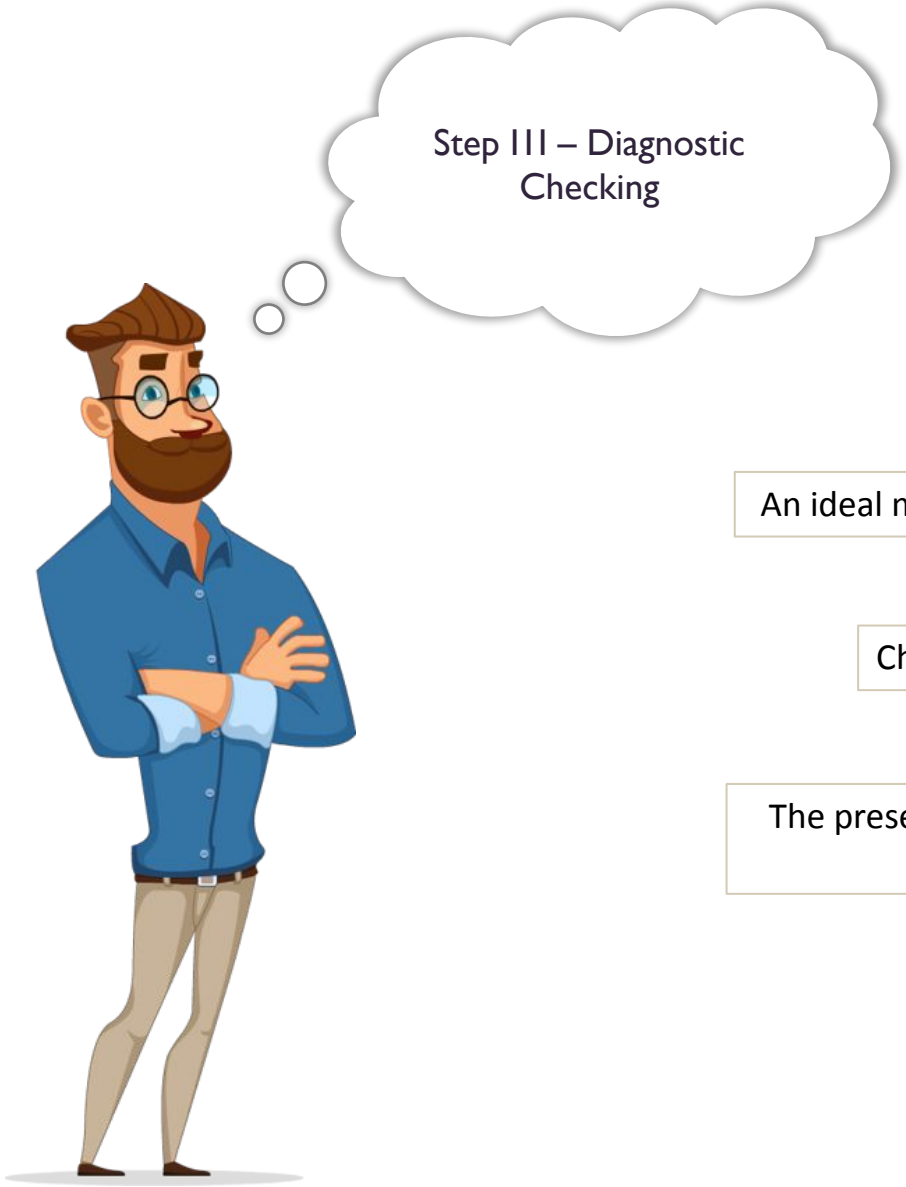
A Gaussian distribution with a mean of zero and a symmetrical variance.

Use density plots, histograms, and Q-Q plots that compare the distribution of errors to the expected distribution.

A non-Gaussian distribution may suggest an opportunity for data pre-processing.

A skew in the distribution or a non-zero mean may suggest a bias in forecasts that may be correct.

Steps to build ARIMA Model



An ideal model would leave no temporal structure in the time series of forecast residuals

Check by creating ACF and PACF plots of the residual error time series.

The presence of serial correlation in the residual errors suggests further opportunity for using this information in the model.

Thank You