ARIMA stands for **Autoregressive Integrated Moving Average** models.

ARIMA is a forecasting technique that projects the future values of a series based entirely on its own **inertia**. **Its main application is in the area of short term forecasting requiring at least 40 historical data points. It works best when your data exhibits a stable or consistent pattern over time with a minimum amount of outliers.** Sometimes called Box-Jenkins (after the original authors), ARIMA is usually superior to exponential smoothing techniques when the data is reasonably long and the correlation between past observations is stable.

**Basic Concepts:**

# The first step in applying ARIMA methodology is to check for stationarity.

# "Stationarity" implies that the series remains at a fairly constant level over time.

# If a trend exists, as in most economic or business applications, then your data is NOT stationary. The data should also show a constant variance in its fluctuations over time. This is easily seen with a series that is heavily seasonal and growing at a faster rate.

# In such a case, the ups and downs in the seasonality will become more dramatic over time. Without these stationary conditions being met, many of the calculations associated with the process cannot be computed.

**Differencing:**

**If a graphical plot of the data indicates nonstationarity, then you should "difference" the series.**

**Differencing is an excellent way of transforming a nonstationary series to a stationary one.**

**This is done by subtracting the observation in the current period from the previous one. If this transformation is done only once to a series, you say that the data has been "first differenced". This process essentially eliminates the trend if your series is growing at a fairly constant rate. If it is growing at an increasing rate, you can apply the same procedure and difference the data again. Your data would then be "second differenced".**

**Autocorrelations:**

**"Autocorrelations" are numerical values that indicate how a data series is related to itself over time. More precisely, it measures how strongly data values at a specified number of periods apart are correlated to each other over time. The number of periods apart is usually called the "lag". For example, an autocorrelation at lag 1 measures how values 1 period apart are correlated to one another throughout the series. An autocorrelation at lag 2 measures how the data two periods apart are correlated throughout the series. Autocorrelations may range from +1 to -1. A value close to +1 indicates a high positive correlation while a value close to -1 implies a high negative correlation. These measures are most often evaluated through graphical plots called "correlagrams". A correlagram plots the auto- correlation values for a given series at different lags. This is referred to as the "autocorrelation function" and is very important in the ARIMA method.**

Autoregressive Models:

**ARIMA methodology attempts to describe the movements in a stationary time series as a function of what are called "autoregressive and moving average" parameters. These are referred to as AR parameters (autoregessive) and MA parameters (moving averages). An AR model with only 1 parameter may be written as...**

**X(t) = A(1) \* X(t-1) + E(t)**

**where X(t) = time series under investigation**

**A(1) = the autoregressive parameter of order 1**

**X(t-1) = the time series lagged 1 period**

**E(t) = the error term of the model**

**This simply means that any given value X(t) can be explained by some function of its previous value, X(t-1), plus some unexplainable random error, E(t). If the estimated value of A(1) was .30, then the current value of the series would be related to 30% of its value 1 period ago. Of course, the series could be related to more than just one past value. For example,**

**X(t) = A(1) \* X(t-1) + A(2) \* X(t-2) + E(t)**

**This indicates that the current value of the series is a combination of the two immediately preceding values, X(t-1) and X(t-2), plus some random error E(t). Our model is now an autoregressive model of order 2.**

**Moving Average Models:**

**A second type of Box-Jenkins model is called a "moving average" model. Although these models look very similar to the AR model, the concept behind them is quite different. Moving average parameters relate what happens in period t only to the random errors that occurred in past time periods, i.e. E(t-1), E(t-2), etc. rather than to X(t-1), X(t-2), (Xt-3) as in the autoregressive approaches. A moving average model with one MA term may be written as follows...**

**X(t) = -B(1) \* E(t-1) + E(t)**

**The term B(1) is called an MA of order 1. The negative sign in front of the parameter is used for convention only and is usually printed out auto- matically by most computer programs. The above model simply says that any given value of X(t) is directly related only to the random error in the previous period, E(t-1), and to the current error term, E(t). As in the case of autoregressive models, the moving average models can be extended to higher order structures covering different combinations and moving average lengths.**

**Mixed Models:**

**ARIMA methodology also allows models to be built that incorporate both autoregressive and moving average parameters together. These models are often referred to as "mixed models". Although this makes for a more complicated forecasting tool, the structure may indeed simulate the series better and produce a more accurate forecast. Pure models imply that the structure consists only of AR or MA parameters - not both.**

**The models developed by this approach are usually called ARIMA models because they use a combination of autoregressive (AR), integration (I) - referring to the reverse process of differencing to produce the forecast, and moving average (MA) operations. An ARIMA model is usually stated as ARIMA(p,d,q). This represents the order of the autoregressive components (p), the number of differencing operators (d), and the highest order of the moving average term. For example, ARIMA(2,1,1) means that you have a second order autoregressive model with a first order moving average component whose series has been differenced once to induce stationarity.**

**Picking the Right Specification:**

**The main problem in classical Box-Jenkins is trying to decide which ARIMA specification to use -i.e. how many AR and / or MA parameters to include. This is what much of Box-Jenkings [1976] was devoted to the "identification process. It depended upon graphical and numerical eval- uation of the sample autocorrelation and partial autocorrelation functions.**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

ARIMA methodology does have its limitations. These models directly rely on past values, and therefore work best on long and stable series. Also note that ARIMA simply approximates historical patterns and therefore does not aim to explain the structure of the underlying data mechanism.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

arima(x, order=c(58), method="CSS")

##### Arguments

**x -** a univariate time series of class ts.

**order -** A specification of the non-seasonal part of the ARIMA model: the three components (p, d, q) are the AR order, the degree of differencing, and the MA order.

p - AR order

d - Degree of Differencing

q- MA order

Conditional sum-of-squares is provided mainly for expositional purposes. This computes the sum of squares of the fitted innovations from observation n.cond on, (where n.cond is at least the maximum lag of an AR term), treating all earlier innovations to be zero. Argument n.cond can be used to allow comparability between different fits. The ‘part log-likelihood’ is the first term, half the log of the estimated mean square. Missing values are allowed, but will cause many of the innovations to be missing.