Semester :VI Subject : DAV Academic Year: 2023 - 2024

ARIMA Model:

An autoregressive integrated moving average (ARIMA) model is a statistical tool utilized for analyzing time series data, aimed at gaining deeper insights into the dataset or forecasting forthcoming trends.

An AR model with lag p is represented as:
yt=C+b1*yt-1+b2*yt-2+ . . +bp*yt-p+E

Where yt) is the observation at time t

yt-1 is the observation just before yt ; C is the constant and E is the error term. So, in simple term yt here is a function of its own lags.

MA it is a function of the lagged errors and not yt itself. The only difference between AR and MA lies in the independent driver. For AR its lagged version of the dependent variable yt and for MA it's the lagged version of the error E

A typical Moving Average equation of lag q would therefore look like:

$yt= \alpha+Et+ \mu 1Et-1+ \mu 2Et-2+ \ . \ . \ . + \mu qEt-q$

An ARIMA model, as stated earlier is the combination of both the AR and MA part.

What is "I" in ARIMA?
The "I" stands for Integrated or the differencing component which is represented by d. Broadly speaking, the value of d determines the degree of difference that is required to make the series stationary.

What is Stationarity?
Stationarity means that the properties of the time series do not change with reference to the time. But that doesn't mean the series is static and not changing. What it means is that the inherent properties of the series, like the central tendency and so on, remain similarly related to changing time.

How do we assess the stationarity?
Majorly there are few ways of assessing whether a series is stationary or not. The most important and widely used one is a statistical measure, which is called the Augmented Dickey-Fuller test or ADF test.

The ADF test is a unit root test, and by the application, the test finds the presence of a unit root, which is equivalent to a non-stationary component in a time series. The unit root is a characteristic of a time series that makes it non-stationary. If the test is able to find the unit root, the ADF test is positive and p value of the test is greater than the significance level.

The hypotheses in case of an ADF test is:
H0 (Null): Presence of Unit Root i.e., series is non-stationary

H1 (Alternate): Unit root is not present; i.e., series is stationary

Semester :VI_____ Subject :_____DAV_____ Academic Year: 2023 - 2024

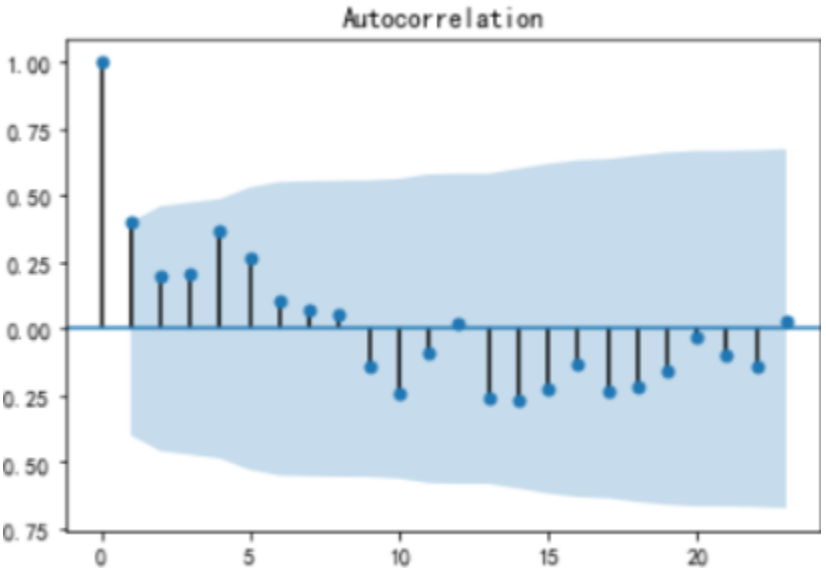
There are standard packages available in python through which ADF tests can be performed.

Upon performing the ADF test and when non-stationarity is found, then comes the application of the integrated part of the ARIMA model. The series is difference with lag terms to make it stationary or, in other words, to get away from the unit root component.

How do we select the degree of differencing required to make the series stationary?

The answer to that lies in the ACF plot or (complete) auto-correlation plot,, which provides the values of auto-correlation of any time series with its lagged values over a series of times. In simple terms, the ACF plot is a wonderful visualization of how the present values of a time series is related to its past values.

The figure below shows what an ACF plot looks like. As can be seen, the blue shaded area is the acceptable zone, and hence for this specific case, first lag onwards, the time series doesn't have the autocorrelation problem.



The ARIMA function from the statsmodels package in python requires three parameters.
p = No. of lags required for the AR Part
q= no. of lags required for the MA part
d = No. of differencing required for reaching a near stationary series using ADF testing, differencing, and ACF plots.

To further simplify, a perfect differencing for a time series may be obtained as the minimum differencing required to get a near-stationary series that roams around a defined mean, and the ACF plot also doesn't have too many numbers of lags outside the acceptable region.