Autoregressive Model Integrated Moving Average (ARIMA)

ARIMA

- It is used in the statistical analysis model that uses the time series data to either better understand the data 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 the past periods.
- Autoregression (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 the time series to become stationary
- Moving average (MA): incorporates the dependency between an observation and a residual error from a moving average model applied to lagged observations.

Assumptions

ARIMA models work on the assumption of stationarity (i.e. they must have a constant variance and mean). If your model is non-stationary, you'll need to transform it before you can use ARIMA.

Keys Points

- Autoregressive integrated moving average (ARIMA) models predict future values based on past values.
- ARIMA makes use of lagged moving averages to smooth time series data.
- They are widely used in technical analysis to forecast future security prices.
- Autoregressive models implicitly assume that the future will resemble the past.
- Therefore, they can prove inaccurate under certain market conditions, such as financial crises or periods of rapid technological change.

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.
- The parameters can be defined as:
  - p: the number of lag observations in the model; also known as the lag order.
  - d: the number of times that the raw observations are differenced; also known as the degree of differencing.
  - q: the size of the moving average window; also known as the order of the moving average.
- The parameters take the value of integers and must be defined for the model to work. They can also take a value of 0, implying that they will not be used in the model. In such a way, the ARIMA model can be turned into:
  - ARMA model (no stationary data, d = 0)
  - AR model (no moving averages or stationary data, just an autoregression on past values, d = 0, q = 0)
  - MA model (a moving average model with no autoregression or stationary data, p = 0, d = 0)

Therefore, ARIMA models may be defined as:

- ARIMA(1, 0, 0) – known as the first-order autoregressive model
- ARIMA(0, 1, 0) – known as the random walk model
- ARIMA(1, 1, 0) – known as the differenced first-order autoregressive model

Steps in ARIMA

1. Prepare your data by using transformations (e.g. square roots or logarithms) to stabilize the variance and differencing to remove remaining seasonality or other trends.
2. Identify any processes that appear to be a good fit for your data.
3. Find which model coefficients provide the best fit for your data. This step is computationally complex and usually performed by a computer. Akaike's Information Criterion (AIC) is one option: if you compare two models, the one with the lower AIC is usually the "better" model.
4. Test the models' assumptions to see how well the model holds up to closer scrutiny. If your chosen model is inadequate, repeat steps 2 and 3 to find a potentially better model.
5. Compute forecasts on your chosen model with computer software.

Applications

1. Forecasting the quantity of a good needed for the next time period based on historical data.
2. Forecasting sales and interpreting seasonal changes in sales
3. Estimating the impact of marketing events, new product launches, and so on.

Limitations

1. Although ARIMA models can be highly accurate and reliable under the appropriate conditions and data availability, one of the key limitations of the model is that the parameters (p, d, q) need to be manually defined; therefore, finding the most accurate fit can be a long trial-and-error process.
2. Similarly, the model depends highly on the reliability of historical data and the differencing of the data. It is important to ensure that data was collected accurately and over a long period of time so that the model provides accurate results and forecasts.

Important Questions

1) What is ARIMA model used for?

It's a model used in statistics and econometrics to measure events that happen over a period of time. The model is used to understand past data or predict future data in a series

2) How many observations are needed for ARIMA?

For autoregressive integrated moving average (ARIMA) models, the rule of thumb is that you should have at least 50 but preferably more than 100 observations (Box and Tiao 1975).

3) Is stationarity required for ARIMA?

If a time series has a trend or seasonality component, it must be made stationary before we can use ARIMA to forecast