**Time series forecasting of returns**

Submitted By

Parul Moriya (18HS20025)

Aim: To predict returns on a stock of Piramal enterprises ltd. using ARIMA forecasting method.

**Design and methodology**

Autoregressive Integrated Moving Average (ARIMA) Model:

ARIMA stands for Autoregressive Integrated Moving Average. ARIMA is also known as Box-Jenkins approach. Box and Jenkins claimed that non-stationary data can be made stationary by differencing the series, Yt. The general model for Yt is written as,

Yt =ϕ1Yt−1+ϕ2Yt−2…ϕpYt−p +ϵt + θ1ϵt−1+ θ2ϵt−2 +…θqϵt−q

Where Yt is the differenced time series value, ϕ and θ are unknown parameters and ϵ are independent identically distributed error terms with zero mean. Here,Yt is expressed in terms of its past values and the current and past values of error terms.

The ARIMA model combines three basic methods:

* AutoRegression (AR) – In auto-regression, the values of a given time series data are regressed on their own lagged values, which is indicated by the “p” value in the ARIMA model.
* Differencing (I-for Integrated) – This involves differencing the time series data to remove the trend and convert a non-stationary time series to a stationary one. This is indicated by the “d” value in the ARIMA model. If d = 1, it looks at the difference between two-time series entries, if d = 2 it looks at the differences of the differences obtained at d =1, and so forth.
* Moving Average (MA) – The moving average nature of the ARIMA model is represented by the “q” value which is the number of lagged values of the error term.

This model is called Autoregressive Integrated Moving Average or ARIMA(p,d,q) of Yt.

**Data used for Return Analysis**

The data of last 5 years for of Piramal enterprises ltd have been used, daily open and close prices were gathered from the website of Yahoo Finance. For statistical analysis, daily returns are calculated by using the following formula

Rt= (Pt – Pt-1)/Pt-1 ∗ 100

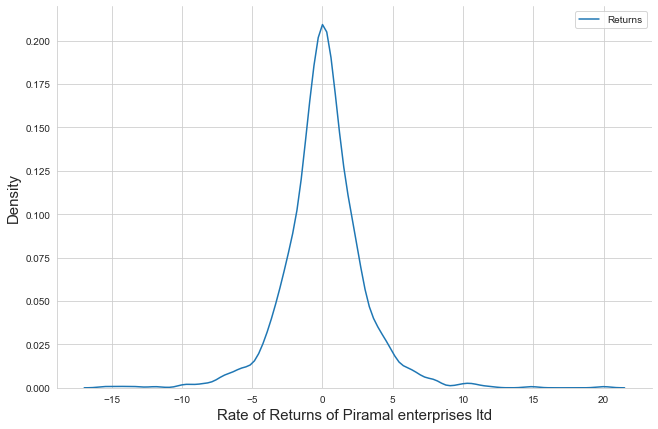
Where,

Rt = Return of Index

Pt  = Closing price of the day

Pt-1 = Closing price of the previous day

**Distribution of Rate of Returns**



* The Rate of Returns are distributed approximately between -10 and 10.
* The Expected Rate of Return is 1.940286
* The Standard Deviation or the Risk associated with such Returns is 7.93148

**Testing For Stationarity of the Data**

Applied ADF test on the returns series data to check for stationarity

Ho: It is non stationary

H1: It is stationary

ADF Test Statistic : -22.348266972843557

p-value : 0.0

Lags Used : 1

Number of Observations Used : 1229

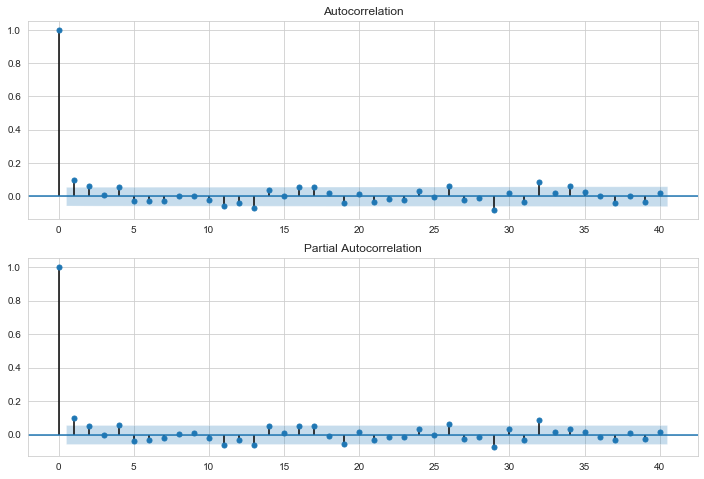
strong evidence against the null hypothesis(Ho), reject the null hypothesis. Data has no unit root and is stationary

* The p-value of 0.0 from the ADF test tells us that the series is stationary. If the series were to be non-stationary, we would have first differenced the returns series to make it stationary.

## **Autocorrelation and Partial Autocorrelation**

* Identification of an AR model is often best done with the PACF.
  + For an AR model, the theoretical PACF “shuts off” past the order of the model. The phrase “shuts off” means that in theory the partial autocorrelations are equal to 0 beyond that point. Put another way, the number of non-zero partial autocorrelations gives the order of the AR model. By the “order of the model” we mean the most extreme lag of x that is used as a predictor.
* Identification of an MA model is often best done with the ACF rather than the PACF.
  + For an MA model, the theoretical PACF does not shut off, but instead tapers toward 0 in some manner. A clearer pattern for an MA model is in the ACF. The ACF will have non-zero autocorrelations only at lags involved in the model.

p,d,q p AR model lags d differencing q MA lags



From these plots I have select AR order = 1 and MA order = 0. Thus, the ARIMA parameters will be (2,0,2).

**Appling Model**

|  |  |  |  |
| --- | --- | --- | --- |
| ARMA Model Results | | | |
| Dep. Variable: | Returns | No. Observations: | 1231 |
| Model: | ARMA(1, 0) | Log Likelihood | -3014.872 |
| Method: | css-mle | S.D. of innovations | 2.802 |
| Date: | Sun, 28 Feb 2021 | AIC | 6035.744 |
| Time: | 13:38:23 | BIC | 6051.091 |
| Sample: | 0 | HQIC | 6041.518 |
|  |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | coef | std err | z | P>|z| | [0.025 | 0.975] |
| const | 0.0981 | 0.089 | 1.108 | 0.268 | -0.075 | 0.272 |
| ar.L1.Returns | 0.0982 | 0.028 | 3.456 | 0.001 | 0.043 | 0.154 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Roots | | | | |
|  | Real | Imaginary | Modulus | Frequency |
| AR.1 | 10.1808 | +0.0000j | 10.1808 | 0.0000 |

**Interpreting the Results**

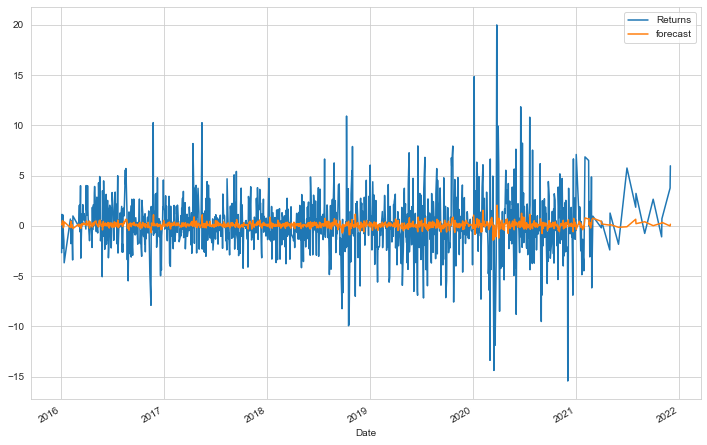
* The p-value of the order-1 AR comes out to be 0.001, which means it is statistically significant.

Estimation through Auto ARIMA Approach

Different ARIMA models are estimated using Akaike Information Criteria (AIC). AIC is used to determine the model best fits a set of data series and it choose the best model to forecast the future data. This is based upon the estimated log-likelihood of the model, number of observations and number of parameters in the model. By using ARIMA models, the number of Auto Regressive Moving Average (ARMA) terms could be determined. The maximum number of Auto Regressive (AR) or Moving Average (MA) coefficients has been specified to determine the number of ARMA 7 International Journal of Pure and Applied Mathematics Special Issue terms, then to estimate every model up to those maxima and then each model could be evaluated using its information criterion. After estimating each model along with calculated criterion, the model could be chosen based on lowest AIC.

* The AIC value is coming out be least for AR(1) , suggests that the chosen values are correct.

**Rate of returns based on the forecasting**



**Conclusion**

* To conclude, in this analysis the ARIMA model is been applied to forecasting stock price returns using python programming language.
* Forecasting with Auto ARIMA provides a prediction based on historical data, in which data has been applied by first order difference to remove white noise problems.
* In validation, the forecasted values are compared with actual values.