FINANCIAL ECONOMETRICS ASSIGNMENT

UNIVARIATE TIME-SERIES ANALYSIS

JAWAHARLAL NEHRU UNIVERSITY

Submitted by:
Nandita Kharbanda

Introduction

This project aims to forecast the future returns of ICICI Bank Ltd. using the Box-Jenkins method of univariate time series analysis. This empirical analysis employs the time series data from the NSE stock prices of ICICI Bank Ltd. which is one of the largest private sector banks in India, with a strong presence in the retail and corporate banking segments. As a publicly listed company, ICICI Bank Ltd.'s performance is closely watched by investors, financial analysts, and other stakeholders. One important aspect of the bank's performance is its returns, which are a measure of the profitability of its operations.

The total number of observations included in the data is 1234 days, and the timeframe chosen is from April 10, 2018 to April 5, 2023. Entire analysis is done of STATA.

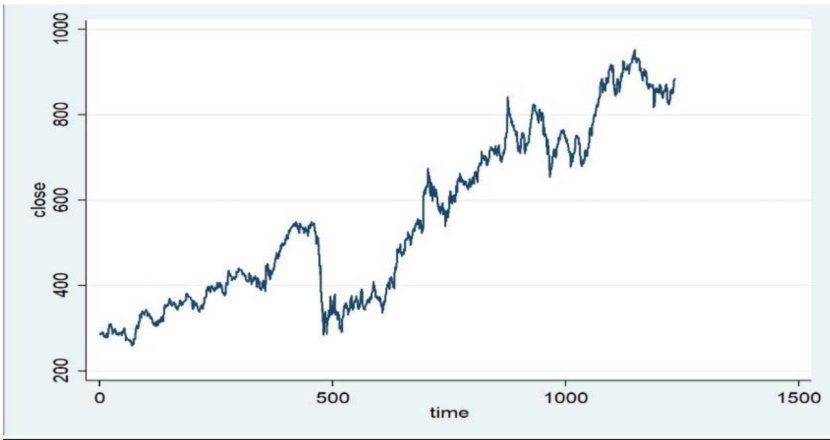
This project aims to analyze the time dynamics of the returns of ICICI Bank Ltd. using the Box-Jenkins method of univariate time series analysis. By testing for weak-form efficiency, fitting an appropriate model to the data, and generating forecasts for the future returns, this project can provide valuable insights for investors and financial analysts.

Below is the descriptive statistics of the closing prices

Variable	Obs	Mean	Std. Dev.	Min	Max
close	1,234	552.5686	200.6594	259.25	952.9

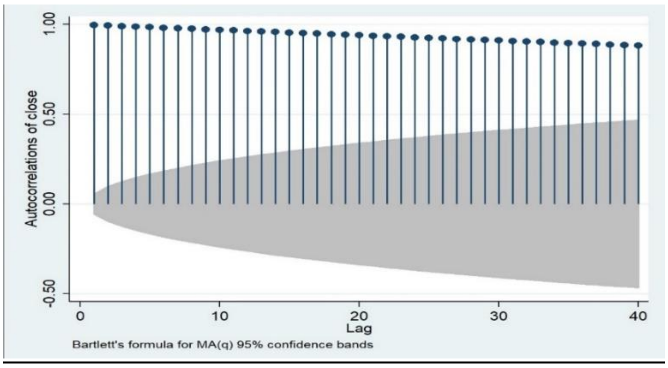
Testing for Stationarity

One common method to check the stationarity of prices is by performing a visual inspection of the price series over time. A stationary time series is one whose statistical properties such as mean and variance remain constant over time.



This graph clearly shows that our prices are non-stationary as there is a consistent upward trend in the prices over time. The upward trend may suggest that the asset is becoming more valuable over time.

I also look for non-stationarity by graphing the correlogram of the autocorrelation function (ACF).



This indicates non-stationarity as there are significant autocorrelations at all the lags.

As correlogram is not sufficient to identify the presence of non-stationarity for that I have performed the Dickey Fuller test.

The Dickey-Fuller test is a statistical test used to determine whether a time series is stationary or non-stationary. The test has a null hypothesis and an alternative hypothesis as follows:

H0: The time series has a unit root, which means that the series is non-stationary.

H1: The time series does not have a unit root, which means that the series is stationary.

```
. dfuller close

Dickey-Fuller test for unit root                   Number of obs   =      1233

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -0.649            -3.430            -2.860            -2.570
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.8594
```

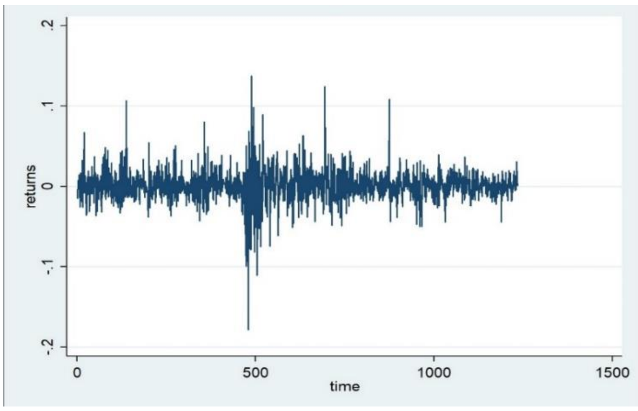
As the p-value is 0.8594 which is greater than 0.05 level of significance. I failed to reject the null hypothesis that the time series has a unit root, which means that the series is non-stationary.

To correct for non-stationarity, I have used returns series instead of prices because returns are more stationary than prices.

I have calculated returns using the given formula on STATA:

Returns = D.close/L.close

To check for stationarity of the returns series, I have plotted the line graph which clearly indicates stationarity as mean is constant over time and the series is not showing any trend.



Next, I also conduct the Dickey Fuller Test on returns

```
. dfuller returns

Dickey-Fuller test for unit root                   Number of obs   =      1232

                               ---------- Interpolated Dickey-Fuller ----------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)            -36.618           -3.430            -2.860            -2.570
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000
```

As p value is 0.0000 which is less than 0.05 level of significance, we reject the null hypothesis and conclude that our returns series is stationary.

<u>**Testing for Weak Form Efficiency**</u>

The weak form efficiency of a financial market is determined by the extent to which past returns can be used to predict future returns. In other words, if a market is weak form efficient, then historical returns should not be able to predict future returns. That is, the returns follow a random process.

H_0: The individual stock prices follow a random process

H_1: The individual stock prices do not follow a random process

In other words;

H_0: The returns are efficient in weak form of efficient market hypothesis (White Noise)

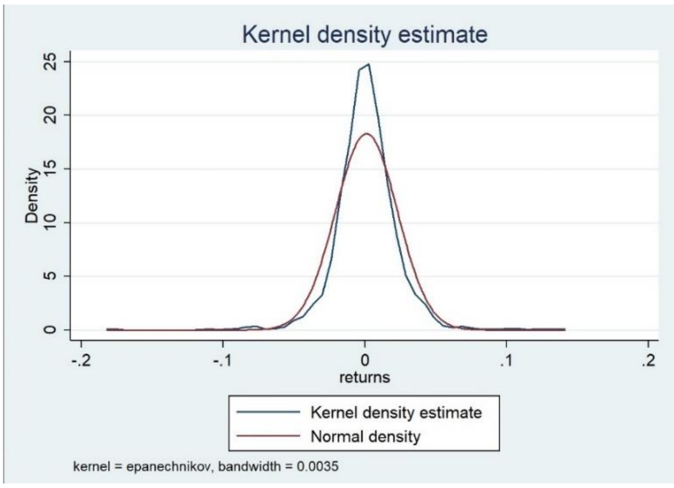
H_1: The returns are not efficient in weak form of efficient market hypothesis (Not White Noise)

To test for weak form efficiency, I have performed the Ljung-Box Test

```
Portmanteau test for white noise
---------------------------------------
 Portmanteau (Q) statistic =    88.3452
 Prob > chi2(40)           =     0.0000
```

The Ljung-Box test is a portmanteau test which can be used to test for randomness in a time series by checking whether the autocorrelations of the series are significant. If the autocorrelations are significant, it suggests that there are patterns or dependencies in the data that are not random. As we can see the results of the Portmanteau (Q) statistic are significant, as p-value is 0.0000 which is less than 0.05 level of significance. Therefore, we can reject the null hypothesis and conclude that the returns are weak form inefficient and they do not follow a random process.

Next, I have plotted a Kernel Density Graph below



The Kernel Density plot of returns appears to be leptokurtic, showing a more peaked and thicker-tailed distribution compared to a normal distribution.

Jarque-Bera Test

```
Jarque-Bera normality test:   3658 Chi(2)      0
Jarque-Bera test for Ho: normality:
```

The Jarque-Bera normality test is conducted to assess whether a given sample of data follows a normal distribution. In this case, the test statistic is 3658, and the associated p-value is 0.

Interpreting the results, the null hypothesis (Ho) states that the data follows a normal distribution. With a p-value of 0, which is less than any conventional significance level (e.g., 0.05), we reject the null hypothesis. This suggests that the data does not follow a normal distribution.

Therefore, based on the Jarque-Bera test, we can conclude that the data does not exhibit a normal distribution.

Box-Jenkins Approach (METHODOLOGY)

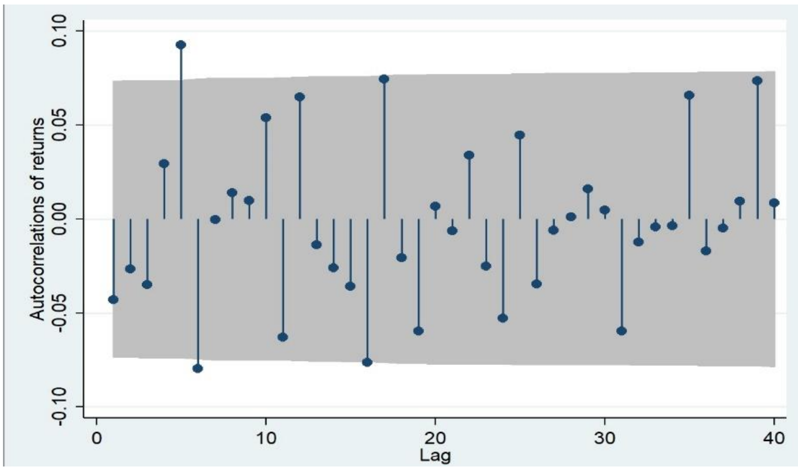
Now that we have established the stationarity of our return series, we can proceed with applying the Box-Jenkins methodology to identify and fit a suitable time series model. This approach allows us to analyze the temporal dynamics of the data and make predictions for future values based on the identified model.

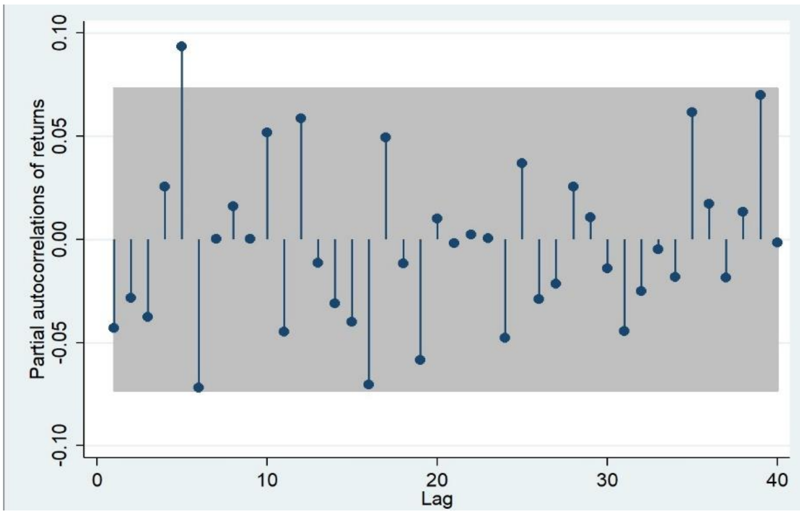
1.IDENTIFICATION

Here, we identify the parameters of an ARMA model for the data. Two diagnostic plots that can be used to help choose the *p* and *q* parameters of the ARMA or ARIMA are **Autocorrelation Function (ACF)** and **Partial Autocorrelation Function (PACF).** Thus, we plot AC and PAC of returns.

- Autocorrelation Function (ACF) determines "q"
- Partial Autocorrelation (PACF) determined "p"
- d is the degree of differencing.

As our returns series is stationary, degree of differencing is 0, while p and q are determined using the correlogram of ACF and PACF of the returns series.





ACF of returns has a significant lag at lag 5, 6. PACF of returns has significant lag at 5. As they lie outside the confidence interval.

I have also set up a loop to estimate all possible combinations of ARIMA(p,d,q) models with p ranging from 0 to 8, q ranging from 0 to 8, and d set to 0.

Based on the loop I have generated and considering the significant lags of ACF and PACF I have temporarily selected the following models for estimation.

ARIMA(5,0,5), ARIMA(5,0,6), ARIMA(6,0,5), ARIMA(1,0,1), ARIMA(2,0,1), ARIMA(2,0,2), ARIMA(3,0,2), ARIMA(4,0,1), ARIMA(5,0,2), ARIMA(6,0,0), ARIMA(6,0,8).

2. ESTIMATION

In this step we will estimate our possible models which we have identified above. Our objective is to select a model which is parsimonious with most significant coefficients, lowest volatility (SigmaSQ), maximum likelihood statistic and minimum AIC and BIC.

	Log-likelihood	Sigma(Volatility)	AIC	BIC	No. of significant coefficients.
ARIMA(5,0,5)	2987.993	.0214425	-5951.985	-5890.579	0
ARIMA(5,0,6)	2991.843	.0213562	-5959.686	-5898.28	10
ARIMA(6,0,5)	2988.002	.0214422	-5950.003	-5883.479	0
ARIMA(1,0,1)	2970.169	.0217567	-5932.338	-5805.869	0
ARIMA(2,0,1)	2970.332	.0217541	-5930.664	-5802.078	0
ARIMA(2,0,2)	2975.426	.0216631	-5938.853	-5811.864	4
ARIMA(3,0,2)	2977.858	.0216234	-5941.715	-5818.365	5
ARIMA(4,0,1)	2972.431	.0217168	-5930.861	-5820.635	1
ARIMA(5,0,2)	2979.973	.0215838	-5941.946	-5856.937	4
ARIMA(6,0,0)	2979.957	.0215733	-5943.914	-5880.345	3
ARIMA(6,0,8)	2993.099	.0213328	-5954.197	-5872.382	1

Based on the above table, ARIMA(5,0,6) is found to be the most appropriate model with lowest AIC, lowest BIC and highest no. of significant lags. Although, the log-likelihood is not the maximum in ARIMA(5,0,6), it is the highest in ARIMA(6,0,8).

We will choose ARIMA(5,0,6) as it satisfies most of the model selection criteria.

ARIMA Regression results are shown below:

```
ARIMA regression

Sample:  2 - 1234                               Number of obs     =       1233
                                                Wald chi2(10)     =    4287.48
Log likelihood =  2991.843                      Prob > chi2       =     0.0000

------------------------------------------------------------------------------
             |                 OPG
     returns |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
returns      |
       _cons |    .001211   .0001516     7.99   0.000     .0009138    .0015081
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |  -.3262195   .0458751    -7.11   0.000     -.416133    -.236306
         L2. |   -.266183   .0523573    -5.08   0.000    -.3688015   -.1635644
         L3. |   .3577893   .0384349     9.31   0.000     .2824582    .4331204
         L4. |   .3199746   .0451605     7.09   0.000     .2314617    .4084875
         L5. |   .8497558   .0508306    16.72   0.000     .7501296     .949382
             |
          ma |
         L1. |   .2898606    .063425     4.57   0.000     .1655499    .4141714
         L2. |    .243897          .        .       .            .           .
         L3. |  -.4424171   .0710958    -6.22   0.000    -.5817623   -.3030719
         L4. |  -.2863147   .0940655    -3.04   0.002    -.4706798   -.1019497
         L5. |  -.8121302   .0964209    -8.42   0.000    -1.001112   -.6231487
         L6. |   .0071045    .024397     0.29   0.771    -.0407127    .0549217
-------------+----------------------------------------------------------------
      /sigma |   .0213562   .0005578    38.28   0.000     .0202628    .0224495
------------------------------------------------------------------------------
Note: The test of the variance against zero is one sided, and the two-sided
      confidence interval is truncated at zero.
```

Estat ic results:

```
Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |        Obs  ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
           . |      1,233         .   2991.843      12   -5959.686   -5898.28
-----------------------------------------------------------------------------
```

Checking for Stability and Invertibility

The stability condition for the eigenvalues is checked to determine if the autoregressive (AR) parameters satisfy the condition for stability, while the invertibility condition is checked for the moving average (MA) parameters.

Eigenvalue stability condition

Eigenvalue	Modulus
.9892449	.989245
-.7529291 + .6137034*i*	.971357
-.7529291 - .6137034*i*	.971357
.09445509 + .9492021*i*	.95389
.09445509 - .9492021*i*	.95389

All the eigenvalues lie inside the unit circle.
AR parameters satisfy stability condition.

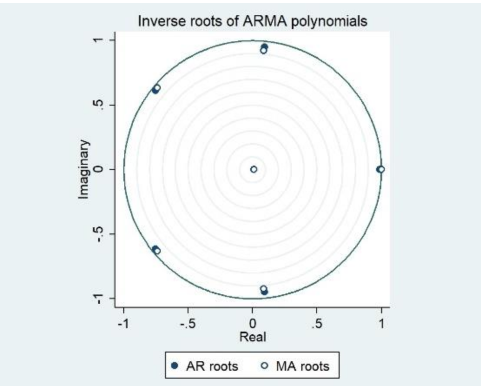
Eigenvalue stability condition

Eigenvalue	Modulus
.9999999	1
-.7381462 + .6330539*i*	.972428
-.7381462 - .6330539*i*	.972428
.08757447 + .923597*i*	.92774
.08757447 - .923597*i*	.92774
.00971079	.009711

All the eigenvalues lie inside the unit circle.
MA parameters satisfy invertibility condition.

For the AR parameters, all the eigenvalues lie inside the unit circle, which indicates that the AR parameters satisfy the stability condition. This implies that the model's autoregressive component is stable and does not lead to explosive or diverging behavior.

Similarly, for the MA parameters, all the eigenvalues also lie inside the unit circle. This indicates that the MA parameters satisfy the invertibility condition, ensuring that the model's moving average component is well-defined and can be inverted to recover the original series.



Based on the eigenvalue stability condition, both the AR and MA parameters of the model satisfy their respective conditions, indicating stability and invertibility of the model.

3. DIAGNOSTIC TESTING

Diagnostic testing is an important step after estimating a statistical model such as an ARIMA model. It involves evaluating the adequacy of the model in describing the observed data by examining the residuals or errors generated by the model.

it is important to ensure that the errors of the estimated model are white noise, which means they are uncorrelated and have constant variance. This is because if the errors exhibit serial correlation or have non-constant variance, it indicates that the model has not fully captured the underlying patterns in the data and there may be further information that could be used to improve the model.

I have estimated the errors of ARIMA(5,0,6) model and then performed the portmanteau test.

Null and Alternate hypothesis of the portmanteau test

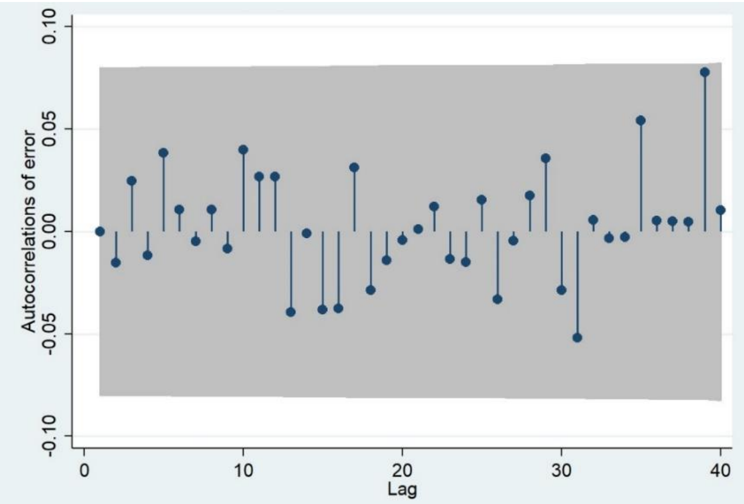
H_0: The errors follow white noise and are not auto-correlated

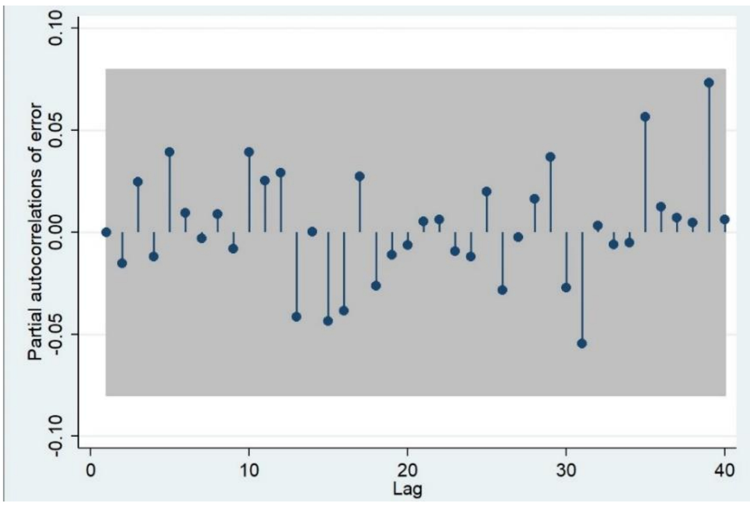
H_1: The errors do not follow white noise and are auto-correlated

```
Portmanteau test for white noise
---------------------------------------
 Portmanteau (Q) statistic =    35.8905
 Prob > chi2(40)           =     0.6558
```

As the p-value is greater than 0.05, we do not reject the null hypothesis and hence we conclude that the errors follow white noise and are not auto-correlated.

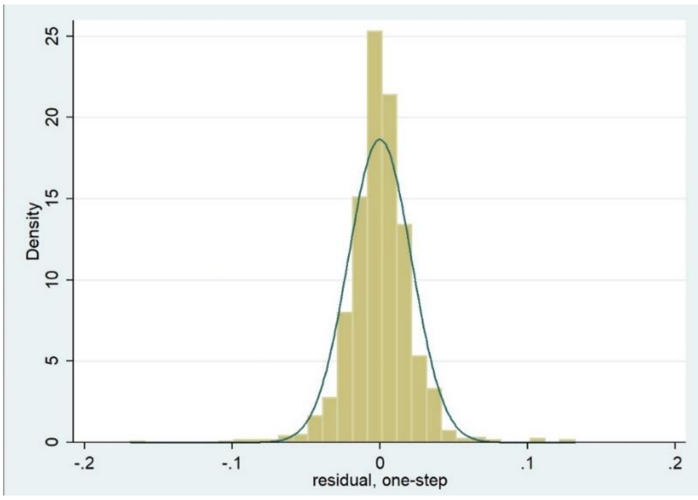
Further I check this through the correlogram of AC and PACF of the residuals.





As all lags of the autocorrelation function (ACF) and partial autocorrelation function (PACF) of the residuals are within the confidence band, it suggests that the residuals are white noise and the model adequately captures the underlying patterns in the data. This indicates that there is no significant autocorrelation left in the residuals after fitting the model.

Below is the histogram of the residuals which shows that the errors follow a normal distribution, it indicates that the residuals are normally distributed around zero, which is a desirable property of a well-fitted model.



I have also performed Dickey-Fuller test to check the stationarity of the errors.

Dickey-Fuller test for unit root — Number of obs = 1232

	Test Statistic	Interpolated Dickey-Fuller: 1% Critical Value	5% Critical Value	10% Critical Value
Z(t)	-35.078	-3.430	-2.860	-2.570

MacKinnon approximate p-value for Z(t) = 0.0000

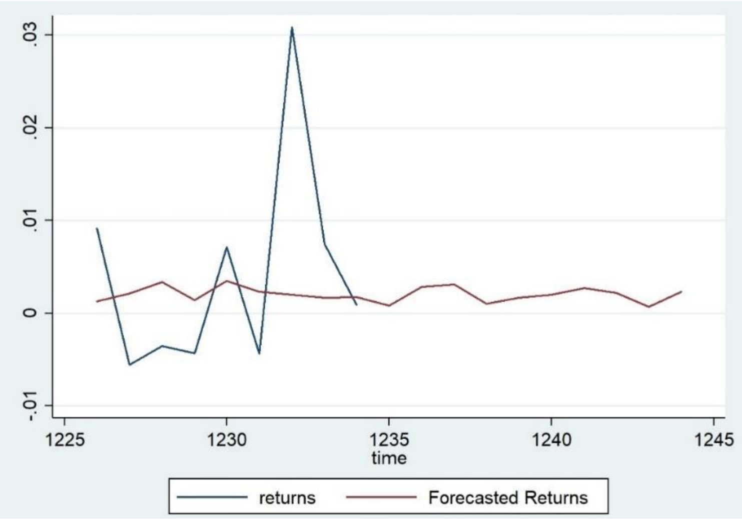
As the p-value is 0.0000 which is less than 0.05, we conclude that our **errors are stationary**.

4. FORECASTING

Here, I will forecast the returns of ICICI Bank Ltd. for the next five period using the fitted model which is ARIMA (5,0,6). The forecast origin is 1235 day and forecasted the returns for the next 5 days.

predict freturns, y dynamic (1235)

Below is the line graph showing actual and forecasted returns



Blue line shows the actual returns and red line shows the forecasted returns. As we can see the above is not a perfect forecast as there may be limitations to the model used for forecasting, which can cause errors in the predicted returns.

Values of the forecasted returns for the next 5 period are shown below

Time	Forecasted Returns
1235	.0008286
1236	.0028432
1237	.0031099
1238	.0009954
1239	.0017058

To evaluate the accuracy of the forecast, the mean-square error (MSE) is calculated. A lower value of MSE indicates a better fit between the actual and forecasted returns. A commonly used benchmark for evaluating the forecast accuracy is the closeness of actual returns to the forecasted returns, which results in a lower value of MSE, indicating a better forecast performance.

Variable	Obs	Mean	Std. Dev.	Min	Max
sfe	1,233	.000457	.0014304	1.76e-13	.0285767

This implies that our forecast is good as the average of the mean square error is close to 0. The analysis of mean-square error indicates that the forecasted returns are in good agreement with the actual returns, thereby validating the suitability of the ARIMA (5,0,6) model for estimating the returns.