Jawaharlal Nehru University

जवाहरलाल नेहरू विश्वविद्यालय

"Time Series Analysis of the Stocks of Tata Power"

Mid Semester Evaluation

Financial Econometrics

Submitted To: -
Prof Mandira Sarma

Submitted By: -
Ritika Priya

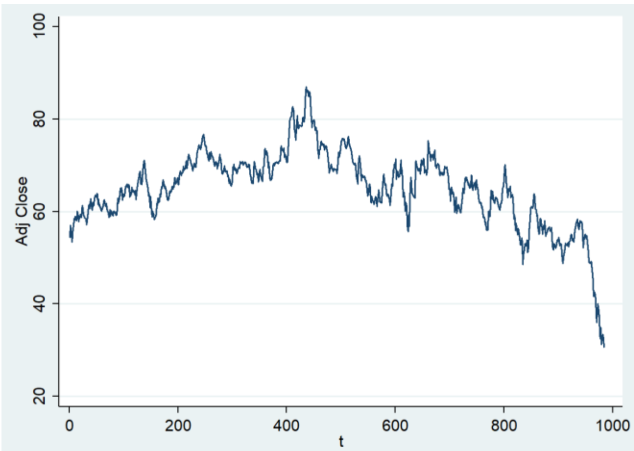
Introduction

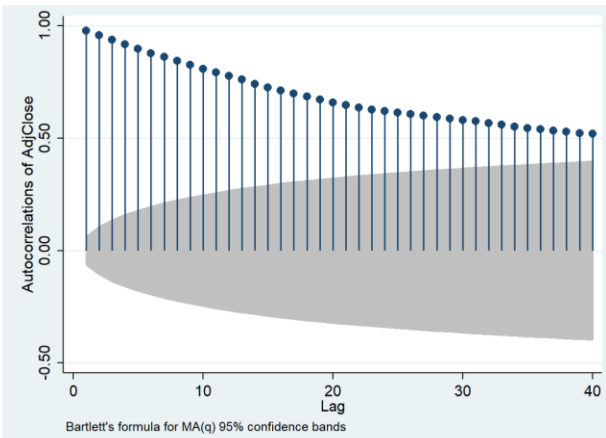
The financial analysis in this paper is done using the daily close price of the stocks of the company Tata Power. Tata Power is India's largest power generation company and a subsidiary of the Tata Group. Daily data is taken for the period of 4 years from 1st April 2016 to 31st March 2020. The pre-covid years are chosen deliberately to make the analysis fair and avoid faulty calculations. The total observations are 984, and all the statistical analysis is done on Stata.

Stationarity of prices

1. Graph: -

Prices are generally non-stationary, and so are the adjusted close prices of the stock, Tata Power. It is clearly visible from this graph as well.
It is clear from the graph that the prices are non-stationery in nature.





In the ACF graph for prices, we can see that all the lags lie outside the confidence interval and hence it also proves non-stationarity of prices.

2. Dickey-fuller test: -

H_0: Process follows a unit root
H_1: Process does not follow a unit root

	Test statistic	1% critical	5% critical	10% critical
z(t)	-0.664	-3.430	-2.860	-2.570

MacKinnon approximate p-value z(t) = 0.8557
The p-value for the D-fuller statistic is 0.92, greater than 0.05. Hence, I do not reject the null hypothesis and conclude that unit root exists in the prices. Therefore, our prices come out to be non-stationary.

So, for this purpose, we work with the returns of these prices. Returns are the dividend payment at time t; there are two kinds of returns-

1. Simple returns (relative returns)
   $R_t = \frac{(P_t+D_t)-P_{t-1}}{P_{t-1}} \times 100$

2. Logarithmic return (relative return)
   $r_t = \log(\frac{P_t+D_t}{P_{t-1}}) \times 100$
   I will be using simple returns in the time series analysis below.

Preliminaries

Before diving into the time series analysis, the first step is to do preliminary data analysis.

1. Summary statistics

Firstly, I have calculated the summary statistics of the returns.

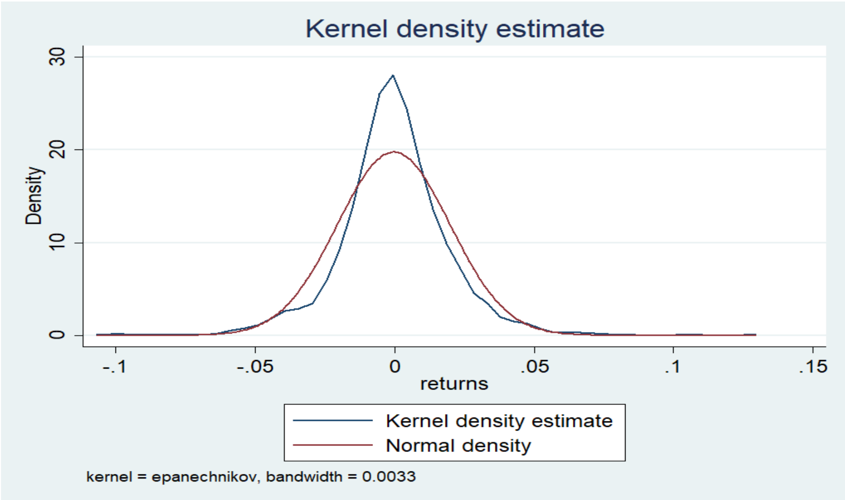
Mean	Std. Dev	Variance	Skewness	Kurtosis
-.00038	.0201	.0004	.1710	7.593

2. Normality test

We need to understand which distribution the data follows.

1. **Graph: -**

   I plot the kernel density graph of the returns to check for normality and can conclude that the data is not normal but leptokurtic, as usually the case with returns is.



2. **J-B test: -**

   The J-B test is a goodness-of-fit test that ascertains whether the data fits the normal distribution.
   H_0: Returns are normally distributed
   H_1: Returns are not normally distributed

p-value	0.0000

   Returns are not normally distributed as the p-value is less than the level of significance.

3. **Shapiro-Wilk test: -**

   This is another famous test used to ascertain the normality of the data. Here, the hypothesis is also same as above

Variable	Obs	Z statistic	P>z
Returns	983	8.656	0.000

   I conclude that the null hypothesis is rejected as the p-value < 0.05 and the returns are not normally distributed.

3. Weak form hypothesis

It suggests that past stock prices or returns do not provide useful information for predicting future returns. Here, I am testing for the weak form hypothesis using the returns of the stock price. The null and alternative hypothesis are as follows:

H_0: The returns are efficient in weak form of efficient market hypothesis.
H_1: The returns are not efficient in weak form of efficient market hypothesis

OR

H_0: Returns follow a random process
H_1: Returns does not follow a random process

I am using Box-Pierce test to test for weak form hypothesis to prove for the weak form hypothesis.

Portmanteau (Q) statistic	**Prob > chi2(40)**
70.2814	0.0022

As p-value < 0.05, the null hypothesis is rejected, and hence returns do not follow a random process and the returns of Tata Powers are weak form inefficient.

Methodology (Box-Jenkins)

The time series analysis will be done using the Box-Jenkins approach. It is a popular forecasting method using regression analysis on time data.
There are 3 significant steps or processes in Box-Jenkins, namely:

1. Identification
2. Estimation Diagnostic checking
3. Forecasting

Let's look at each of these processes in detail.

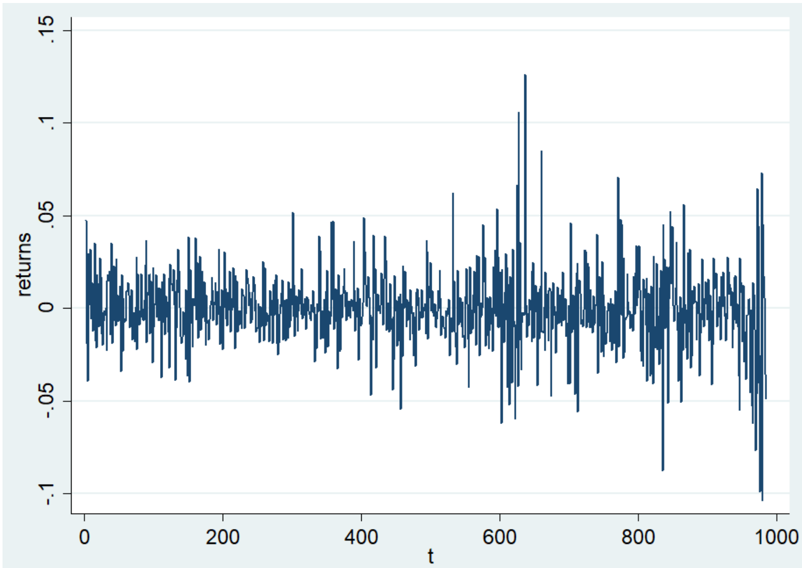
Stage 1: Identification

Identification involves determining the order of the model (p,d,q) to capture the data's dynamic features.
This stage involves two steps:

1. **Checking for stationarity**

1. **Graph: -**

   I plot the returns graph to get a fairer idea of the returns. We can infer from the graph below that returns are indeed stationary.



2. **Augmented Dickey-Fuller test: -**

   H_0: Process follows a unit root
   H_1: Process does not follow a unit root

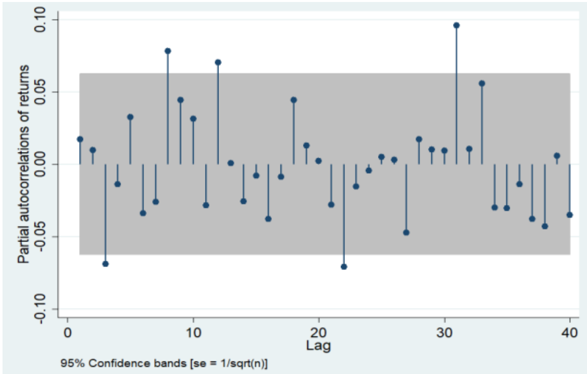
	Test statistic	**1% critical**	**5% critical**	**10% critical**
z(t)	-30.774	-3.430	-2.860	-2.570

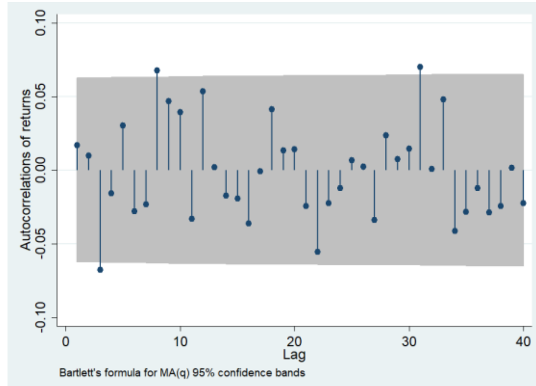
   MacKinnon approximate p-value z(t) = 0.00

   It is clear that the p value< 0.05; I reject the null hypothesis and conclude that the returns are stationary.

2. Determining the model's p, d, & q

As the model is stationary, I don't need to differentiate it further to get the stationary process. Hence, d=0 in thw model.
The p or the AR lags of the model can be predicted by plotting the PACF graphs.
The model's q or MA lags can be expected by plotting the ACF graphs.
The graphs generated from the model are as follows:





From the above two figures, I can infer that at 95% confidence intervals, the significant lags of the model which lie outside the confidence intervals are-

p	3	8	12	22
q	3	8	32	

This gives us a variety of models to check for estimation.

Stage 2: Estimation and Diagnostic checking

This step involves estimating the parameters of the models identified in the previous step. The most appropriate model would be parsimonious and would have white noise errors. I will examine various criteria to decide which ARIMA model is best. These criteria are-

1. **Information criteria: -**
   It provides information on how good an estimated model is while penalising over-parameterized models. The objective is to choose the number of parameters that minimises the information criteria value.
   There are two kinds of information criteria-
   1. AIC (Alkaline Information Criteria)
   2. BIC (Bayesian Information Criteria)

2. **Log-likelihood:** - We will prefer the model which has a higher log-likelihood estimate.
3. **Sigma SQ (volatility):** - Sigma SQ shows the volatility. Hence, we would go for a measure with a low value of Sigma HQ.
4. **Number of significant coefficients: -** The more significant coefficients are, the better the model. Hence, we would select the model which has a higher number of significant coefficients.

I have run almost all possible values for ARIMA (p, 0, q) with a special focus on the lags generated in the previous step. I have shown the most significant models of the whole consignment below.

Models	Sig Coeff	Log-likelihood	Sigma SQ	AIC	BIC
ARIMA(3,0,3)	2	2447.769	.0002262	-4879.538	4840.413
ARIMA(3,0,6)	5	2448.108	.0003028	4874.217	-4820.42
ARIMA(3,0,8)	4	2450.575	.0003119	4875.151	-4811.573
ARIMA(3,0,10)	7	2450.587	.0003225	4879.033	-4805.674
ARIMA(3,0,12)	1	2450.595	.5668448	-4887.11	-4803.969
ARIMA(8,0,3)	3	2248.56	.0002279	-4471.119	-4404.533
ARIMA(8,0,6)	4	2259.763	.9912487	-4489.527	-4412.696
ARIMA(8,0,8)	2	2272.655	.0004219	4507.311	-4409.992
ARIMA(8,0,10)	1	2274.644	.0038021	-4511.288	-4412.569
ARIMA(10,0,3)	5	2454.337	.0003218	-4878.674	-4805.315
ARIMA(10,0,6)	1	2455.21	.696317	-4874.419	-4786.388
ARIMA(5,0,8)	1	2451.654	.0003098	-4873.307	-4799.948
ARIMA(5,0,10)	1	2455.181	.1864198	-4876.361	-4793.221

Parameters	Models
Max significant coefficients	ARIMA(3,0,10)
Max Log-likelihood	ARIMA(8,0,10)
Min SigmaSQ	ARIMA(3,0,3)
Min AIC	ARIMA(8,0,6)
Min BIC	ARIMA(3,0,3)
Most suitable	**ARMIA(3,0,3)**

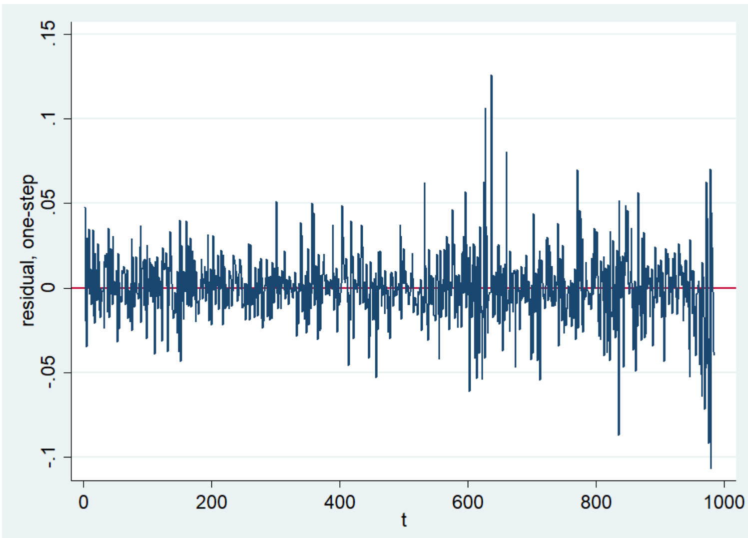
The model ARIMA (3,0,3) generates the most suitable results. Also, its AIC is the second smallest in the entire model. So, I will be doing diagnostic checks using ARIMA (3,0,3).

Diagnostic checking

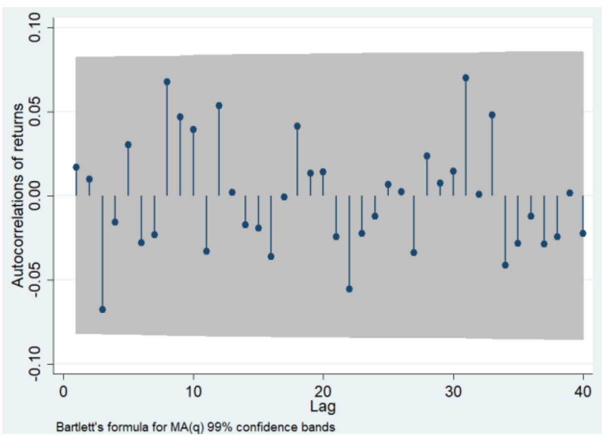
The selected model should be able to generate white noise errors.

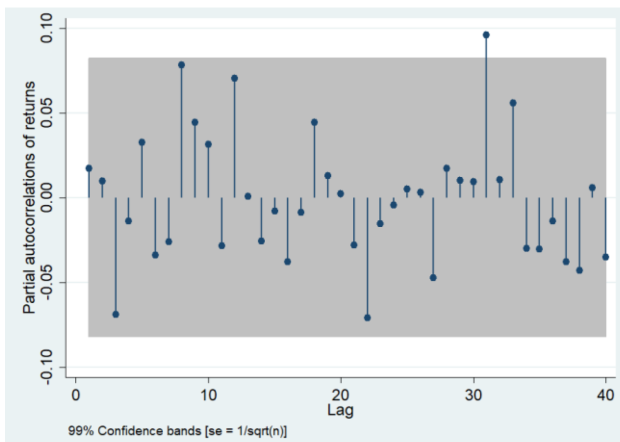
1. **Graphs: -**

   The residuals are plotted against the time t. The residual plot shows a random scatter around zero, with no patterns or trends over time. It suggests that the model is a good fit for the data and the assumptions of the model are met.

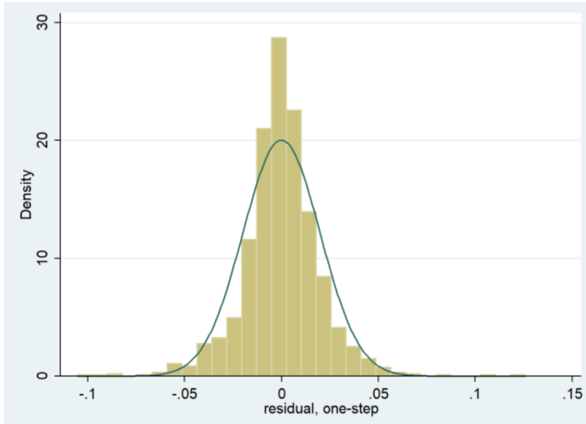


By drawing the correlograms of the error at 99% internal, we can see that there are almost no significant lags hence it means that errors follow a white noise process. It means the model captures almost all information.





A histogram with a mean of 0 is plotted, proving that the residuals follow a normal distribution.



2. Portmanteau tests: -

This statistical test evaluates the model's goodness of fit. It tells whether the residuals have significant autocorrelation indicating the presence of remaining patterns in data not captured by the model.

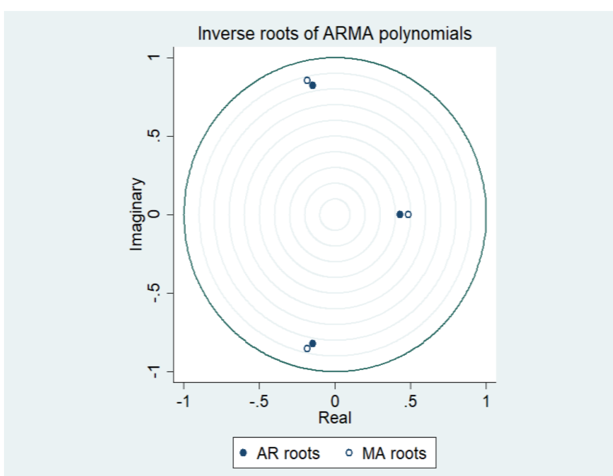
H0: Residuals or errors are not autocorrelated and follow a white noise
H1: Residuals or errors are autocorrelated and do not follow a white noise

Portmanteau statistic	P value
30.9029	0.8486

As p value > 0.05, I cannot reject null alternate hypothesis; hence, residuals are white noise.

3. Condition for stable univariate process: -

We have learnt in class that for the ARMA process to be stationary, the roots of the AR and MA processes should lie inside the unit circle.



We can infer from the graph that the roots lie within the circle, confirming that residuals are white noise and the process is stable.

Step 3: Forecasting

This is the last step, where I forecast the data for the successive 10 time periods or days. The forecasted returns for the last 10 days of our selected period are given in the following table.

Date	t	Return	Forecasted return
3/17/2020	975	-.0370828	-.0058046
3/18/2020	976	-.0988447	-.00687821
3/19/2020	977	-.0256409	.0041811
3/20/2020	978	.07309941	.0026523
3/23/2020	979	-.1035422	.00287333
3/24/2020	980	.0455927	.00078276
3/25/2020	981	.0174417	-.00679679
3/26/2020	982	.0057143	.00763939
3/27/2020	983	-.0355112	.00190346
3/30/2020	984	-.048601	-.0090049

I have forecasted returns for the next 10 periods, too, i.e., from t = 985 and 10 days onward, presented in the table below.

Date	t	Forecasted returns
4/1/2020	985	-.00162581
4/2/2020	986	.0071262
4/3/2020	987	.00157564
4/4/2020	988	-.00476745
4/5/2020	989	.00014793
4/6/2020	990	.00278992

4/7/2020	991	-.00155097
4/8/2020	992	-.00218867
4/9/2020	993	.00099366
4/10/2020	994	.00049846

Now, the graph for the returns and forecasted returns are shown for the past 10 time periods plus the forecasted periods are given below.

We can see that the graph of the actual return and the forecasted return cross each other at 4 time periods. Also, the table and the graph show that the predicted and actual returns are close to each other, proving a good forecast.

To test the forecast's performance, I check for mean square error and see if it is minimum. No, we can see that the mean squared error is close to 0, which means the forecast is good.

Observation	Mean	Std deviation
9	0.0003466	0.0039658

Hence, it can be concluded that the ARIMA(3,0,3) is the best and most suitable model to fit the returns of the 4 years 2016-2020 for the stocks of Tata Power.

Appendix

Summary statistics:

```
.
. summarize returns, detail

                         returns
-------------------------------------------------------------
      Percentiles      Smallest
 1%    -.0546264      -.1035422
 5%    -.0330807      -.0988447
10%     -.021834        -.0875       Obs                 983
25%    -.0100942       -.076655      Sum of Wgt.         983

50%    -.0006948                     Mean          -.0003805
                        Largest      Std. Dev.      .0201852
75%     .0096279       .0730994
90%     .0214621       .0847568      Variance       .0004074
95%       .03132        .106024      Skewness       .1710764
99%     .0536583         .12627      Kurtosis       7.593549
```

Dfuller test for prices:

```
. dfuller AdjClose

Dickey-Fuller test for unit root                   Number of obs   =       983

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -0.664            -3.430            -2.860            -2.570
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.8557
```

Shapiro wilk test:

```
. swilk returns

                   Shapiro-Wilk W test for normal data

    Variable |        Obs       W           V         z       Prob>z
-------------+------------------------------------------------------
     returns |        983    0.94679     33.039     8.656    0.00000

.
```

Augmented Dfuller test for returns:

```
. dfuller returns

Dickey-Fuller test for unit root                   Number of obs   =       982

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)            -30.774            -3.430            -2.860            -2.570
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000
```

Box Pierce test:

```
Portmanteau test for white noise
---------------------------------------
 Portmanteau (Q) statistic =    70.2814
 Prob > chi2(40)           =     0.0022
```

Portmanteau test for error:

```
. wntestq error

Portmanteau test for white noise
---------------------------------------
 Portmanteau (Q) statistic =    30.9029
 Prob > chi2(40)           =     0.8486

.
```