Department of Computer Science & Engineering
SOUTHEAST UNIVERSITY

CSE4000: Research

Dhaka Stock Market Analysis and Prediction with ARIMA Model

A dissertation submitted to the Southeast University in partial fulfillment of the requirements for the degree of B. Sc. in Computer Science & Engineering

Submitted by

Aninda Halder	Sabbir Hasan Apon
ID: 2013100000039	ID: 2013100000057

Supervised by

Ashiqur Rahman
Assistant Professor , Department of CSE
Southeast University

september, 2017

Abstract

Share price prediction is a task where uncertainties and indifferences are very much visible.We are trying to calculate 5 to 6 companies stock data using our method which is called 'ARIMA' model where we represent our thought to approximately predict the prices. Apart from that we use sufficient number of previous exercised data.We will forecast about 30 days future stock data which will be near about approximate.When we were doing our research it clearly reflects that Dhaka Stock Exchange held unsteady equilibrium of data and it cause us various kinds of difficulties.

Contents

List of Figures

Chapter 1

Introduction

Forecasting is becoming an important factor in every fields.A time series is a sequence taken at successive equally spaced points in time.Time series forecasting is the use of a model to predict future values based on previously observed values.

1.1 objective

This study projects an inside view of the application of AR, MA and ARIMA time series model to forecast the Closing prices of some select stocks in the Share market. In this research, we not only devoted our focus on those who are looking to invest in the share market with optimum efficiency, but also upon those who are working on these time series to implement the prediction analysis factor on various different economic fields with the help of ARIMA. We believe, everyone will get benefit to study this research paper where all the algorithms and codes are simply understandable and can be used for a few different purposes.[1]

We set goal to try guideline for the investor when to buy or sell the stocks. This financial instrument has gained a lot of momentum in recent past as Dhaka economy is curbed with factors like changing political scenario, global clues and high inflation etc, so researcher, investors and speculators are in search of different financial instrument to minimize their risk by portfolio diversification. so, it has become necessary to Analysis predict the price.We'll show the results later in this report.

1.2 Motivation

Risk is uncertainties day and age.Its unpredictability has attracted many investors to invest their money on stocks of different companies. This unpredictability has reason to come up with a method to analyze and predict the closing prices of the stocks in the share market.

Forecasting is becoming an important factor in every fields.A time series is a sequence taken at successive equally spaced points in time.Time series forecasting is the use of a model to predict future values based on previously observed values.The previous research documents were good but cant find them proper form wheres there were lots of scrambled data as example some works with just uni variant or multi variant some works without any certain data.however some of the method in gold market some use them rice economic and so on.

1.3 Overviews

This chapter will present a glance about the research topic.though we know share market is hypothetical field and we try to give as much as basic context in our research. And then it will elaborate very easily in step by step process based on motivation and objective.In this research we simplyfied some methods of algorithm.We also try to show some previous work and what they have done before based on this model.After We will find the equation that makes the analysis. Finally, a implementational result of our research will be discussed.

Chapter 2

ARIMA

ARIMA is the integrated version of autoregressive model(AR) and Moving Average(MA).It based on ARMA model. Auto-regressive and Moving Average (ARMA) model is an important method to study time series. This Model converts a non-stationary data to a stationary data before working on it.

2.1 Auto regressive model(AR)

The autoregressive model is used to describe certain time-varying processes in nature, economics, etc. It specifies that the output variable depends linearly on its own previous values and on a stochastic term, thus the model is in the form of a stochastic difference equation.

$$X_t = c + \sum_{i=1}^{p} \phi_i x_{t-i} + \in_i \tag{2.1}$$

2.2 Moving Average(MA)

The moving-average (MA) model is a common approach for modeling univariate time series. The moving-average model specifies that the output variable depends linearly on the current and various past values of a stochastic (imperfectly predictable) term.

$$X_i = c + \sum_{i=1}^{q} \theta_i \in_{t-i} + \in_i \tag{2.2}$$

2.3 Autoregressive Integrated Moving Average (ARIMA)

The AR part of ARIMA indicates that the evolving variable of interest is regressed on its own lagged values. The MA part indicates that the regression error is actually a linear combination of error terms whose values occurred contemporaneously and at various times in the past. The I indicates that the data values have been replaced with the difference between their values and the previous values. The purpose of each of these features is to make the model fit the data as well as possible.

$$Y_i = c + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{i=1}^{q} \theta_i \in_{t-i} + \in_i \tag{2.3}$$

Chapter 3

Theoretical Framework

3.1 Statistical stationarity

If Mean, variance, autocorrelation etc statistical properties are constant over time then a stationary time series become statistical.

Consequently, sets the conditions of its operation such as mean and variance, if they are present, also do not change over time and also whose joint probability distribution does not change when shifted in time is Stationary.

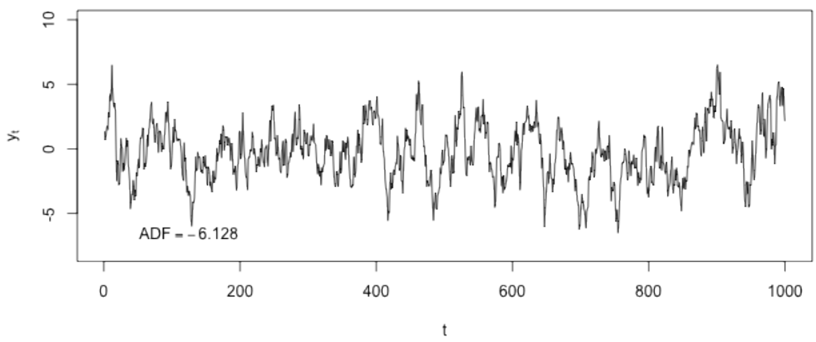

Figure 3.1: Stationary Time Series

For infinite variance a stationary process will not drift too far away from its mean value.
We can simply defined non-stationary that as processes that are not stationary and have statistical properties are deterministic function of time.It is a more complex process then stationary.It necessary before demonstrating make sure that data set is stationary to do so through analysis of the process. [2]

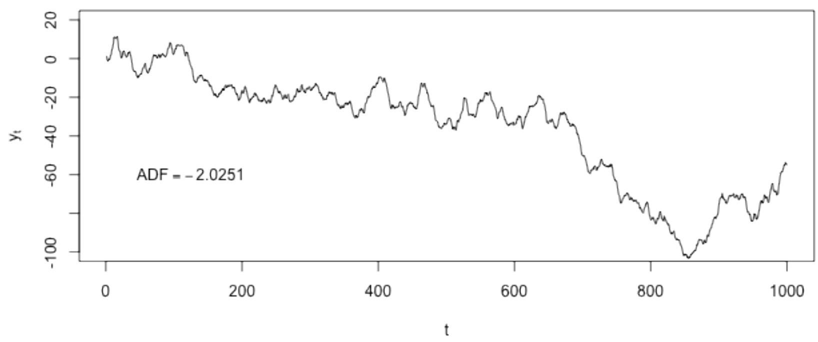

Figure 3.2: Non-Stationary Time Series

3.2 Auto correlation Plot and Partial Autocorrelation Plot

The linear dependence of a variable with itself at two points in time its called autocorrelation.And in general, a partial correlation is a conditional correlation. Assume autocorrelation function and Partial auto correlation function plots determining AR and MA terms.We get MA terms from ACF and AR terms from PACF.

Auto correlation(ACF):

To determine the degree of dependencies in the data ACF plot is used and it represents the autocorrelation dependencies between data. If the ACF plot shows a significant lag at Lag K and there is no significant lag higher than K then K order MA term is needed.

The correlation between X_t and X_{t-1}

$$\rho(X_t, X_{t-1}) = \frac{\sum_{t=2}^{n}(X_t - \bar{X})(X_{t-1} - \bar{X})}{\sum_{t=1}^{n}(X_t - \bar{X})^2} \tag{3.1}$$

Partial Auto correlation(PACF):

If in the PACF plots there is a significant lag at lag K and there is no significant lag higher than K then K order AR term is needed.

The partial autocorrelation of Y_t and Y_{t-k} is the least square regression coefficient of Y_{t-k} when least square is done using Y_t and Y_{t-1} , Y_{t-2} ...... Y_{t-k} terms.

K = 1 : $\mathrm{Y}_t = \alpha_1 + \beta_{11}Y_{t-1} + \zeta$

K = 2 : $\mathrm{Y}_t = \alpha_2 + \beta_{21}Y_{t-1} + \beta_{22}Y_{t-2} + \zeta$

K = 3 : $\mathrm{Y}_t = \alpha_3 + \beta_{31}Y_{t-1} + \beta_{32}Y_{t-2} + \beta_{33}Y_{t-3} + \zeta(3.2)$

partial autocorrelation of order 1,2 and 3 are the parameters β_{11}, β_{22}, β_{33}.

3.3 Standard deviation

Standard deviation is an improvement over mean deviation as a measure of dispersion, and is used most commonly in statistical analysis.It is defined as square root of mean squared deviation

We can also define that average deviation from the mean in a distribution is called standard deviation.

Standard deviation shows that how data is scattered from mean value.[3]

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x - \bar{x})^2}{n-1}} \tag{3.3}$$

3.4 Standard Error

The measure of precision for statistic is standard error. standard error may be define as a statistical term by which accuracy can be measured from a sample that represent a population.

In statistic when a sample proportion deviates from the actual mean or proportion of a population this deviation is the standard error.

3.5 Mean Squared Error (MSE)

Mean Squared Error (MSE) is used to measure the accuracy of a forecasting model. This measure is the average of the sum of the squared differences between the forecast values and the actual time series values.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_t - F_t)^2 \tag{3.4}$$

Where

Observed value $= Y_t$

Predicted value $=F_t$

Error$= (Y_t - F_t)$

3.6 Dikifuller test

Statistically, the Dickey Fuller test tests the null hypothesis of whether a unit root is present in an autoregressive model. The alternative hypothesis is different depending on which version of the test is used, but is usually stationarity or trend-stationarity.

Assume an AR(1) model. The model is non-stationary or a unit root is present if $|\rho|$ =1.

$$\mathrm{y}_t = \rho y_{t-1} + e_t$$
$$\mathrm{y}_t - y_{t-1} = \rho y_{t-1} - y_{t-1} + e_t$$
$$\delta y_t = (\rho - 1) y_{t-1} + e_t =_{t-1} + e_t (3.5)$$

We can estimate the above model and test for the significance of the γ coefficient.
If the null hypothesis is not rejected $\gamma^* = 0$, then y_t is not stationary. Difference the variable and repeat the Dickey fuller test to see if the difference variable is stationary.

3.7 Seasonality

Seasonality can be termed as a type of autocorrelation pattern where the patterns occur every season like monthly,quarterly etc.
For example, in any given year quarterly data have the same pattern in same quarter of that year.
Before a time series model is fitted seasonality must also be corrected.

3.8 White noise

A process is a random of random variables that are uncorrelated, have mean value zero and a finite variance is called white noise.

There are no correlation between its values at different times. It also describes the assumption that each element in a series is a random draw from a population with zero mean and constant variance.

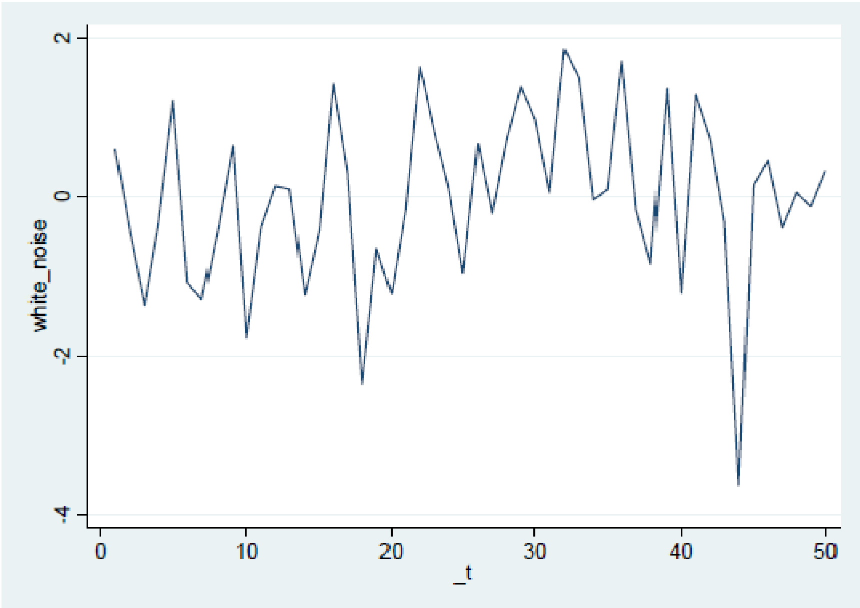

Figure 3.3: white Noise

Autoregressive (AR) and moving average (MA) models correct for violations of this white noise assumption.[4]

Chapter 4

Previous Works

Earlier many people have done their paper work using ARIMA model forecast various types of procedure.Study of Dhaka Stock Exchange-Capital Asset Pricing Model and Efficient Market Hypothesis thesis use ARIMA also applied neural network techniques.. But profound tests shows that DSE dose not hold the actual CAPM.

Stock price prediction Using the ARIMA [5] paper's was applied in NSE NYSE for knowing its efficiency it was tremendous.it got high potential to give a good result in short-term basis.

This [5] paper also define us very well that ARIMA models are known to be robust and efficient in financial time series forecasting especially short term prediction than even the most popular Artificial Neural Network's(ANN) techniques. A Framework for stock Prediction [6] Paper proposes a relatively good framework for stock prediction which can satisfy, but they hope better one should be built when real dataset is obtained

Indian Share Market Forecasting with ARIMA model [7] paper they also tried to forecast th future unknown value of Indian stock market. though it is a common scenario that ups downs are visible in DSE in recent past,that's why we conduct a forecasting measure using ARIMA model.

Chapter 5

Problem

To predict a company share price we will use a univariant time series model. To decrease the error, we will take help from all the theorem that we discuss earlier. Working with share price with a bigger data set is much more challenging , because bigger the data set means higher possibility of randomness which led to poor prediction.

Chapter 6

Implementation

We use AAMRA TECHNOLOGIES LIMITED historical value to predict their share price .We are using historical data from 01/01/2013 to 31/01/2017 . The ploted data are in billow.

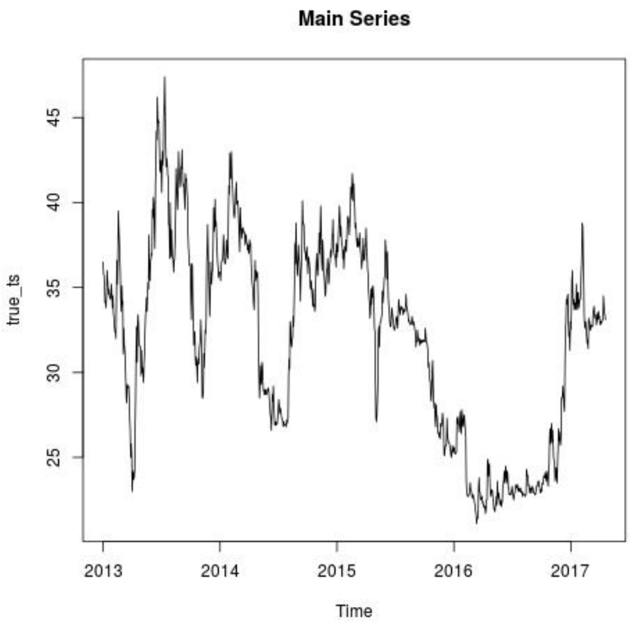

Figure 6.1: Main Serise

We can clearly see that the plotted series is non-stationary,so we did 1st differences to check whether it's stationary or not ,and the plot shows that after doing 1st difference it become stationary .

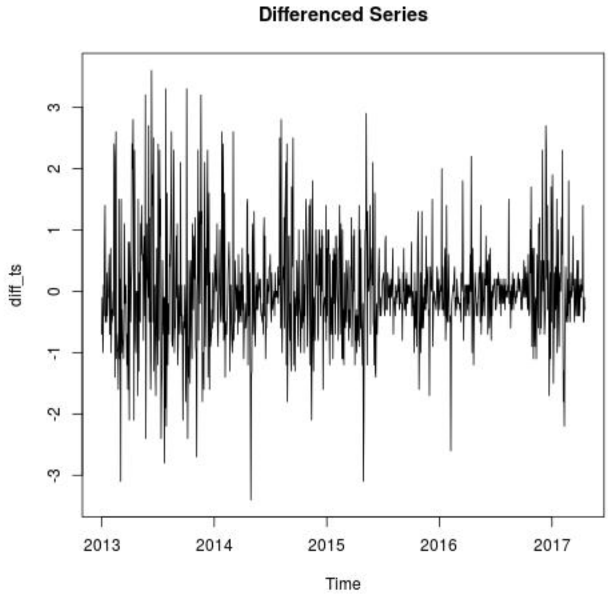

Figure 6.2: 1st Difference Series

In next step we will check white noise of the series . Using Ljung theorem we observed that there is no dependency from lag 1 to 10 .It means the the data are random . We can not go further with a random time series .

Now we will take some other companies to check their white noise. The companies are -

*ACI FORMULATIONS LIMITED

*ANWAR GALVANIZING LIMITED

*APEX FOODS LIMITED

*NATIONAL TEA COMPANY LIMITED

Company Name	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5	Lag 6	Lag 7	Lag 8	Lag 9	Lag 10
AAMRA TECHNOLOGIES LIMITED	0.3285	0.6177	0.6177	0.4667	0.2457	0.2403	0.3121	0.2759	0.1310	0.1509
ACI FORMULATIONS LIMITED	0.08896	0.1782	0.0014	0.0033	0.0028	0.0060	0.0041	0.0035	0.0041	0.0048
ANWAR GALVANIZING LIMITED	0.8727	0.361	0.362	0.0001	0.0002	0.0004	0.0005	0.0010	0.0020	0.0023
APEX FOODS LIMITED	0.0001	0.1000	0.3240	0.129	0.372	0.631	0.287	0.907	0.222	0.301
NATIONAL TEA COMPANY LIMITED	0.6506	0.498	0.0001	0.0005	0.0013	0.0024	0.0050	0.0085	0.0034	0.0040

Table 6.1: white Noise Testing

Among them APEX FOODS LIMITED has the most dependency among all other in every lag . So we took APEX FOODS LIMITED for further analysis . Our goal is to find appropriate ARIMA (seasonal and non-seasonal) model with the help of Auto correlation Function(ACF) and Partial Auto correlation Function(PACF).

Below are the ACF and PACF charts for the first difference values .

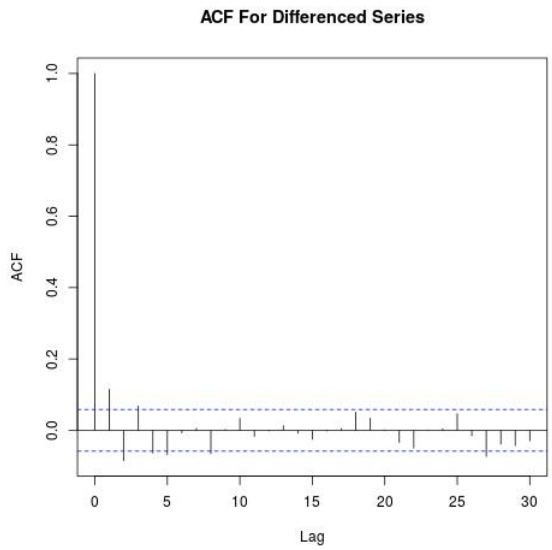

Figure 6.3: Autocorrelation

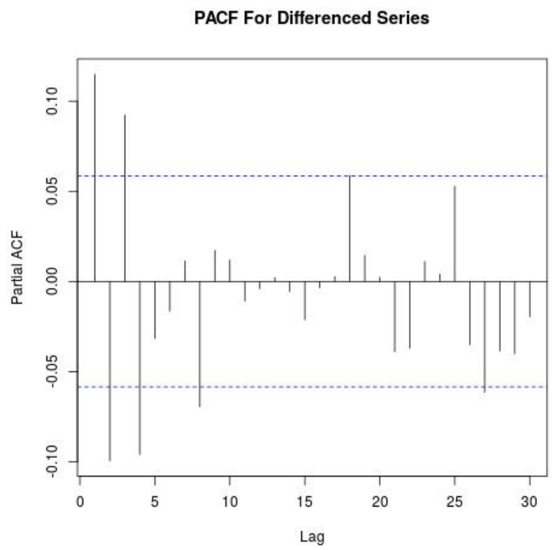

Figure 6.4: Partial Autocorrelation

There are many rules and best practices about how to select the appropriate AR, MA,SAR(Seasonal AR),SMA (Seasonal MA) . From ACF and PACF it is clear that there are some dependency among first 3 lag ,but the graph really don't show any seasonality . Based on that rules we find that there are couple of valid model(without seasonality) for this time series. To find the most optimize model we evaluate all the valid model, and try to find out which one is the most appropriate for this time series .

Chapter 7

Result and Evaluation

Model	AR (1)	AR (2)	AR (3)	MA (1)	MA (2)	MA(3)	AIC
ARIMA (1,1,1)	-0.5862			0.7422			5795.11
ARIMA(2,1,2)	0.3987	0.5495		-0.2575	-0.7135		5790.96
ARIMA(1,1,2)	0.5038			0.6487	-0.0374		5796.45
ARIMA(2,1,3)	-0.5267	-0.5837		0.6766	0.5897	0.1478	5795.7
ARIMA(3,1,2)	-0.4872	-0.0953	0.0458	0.6315	0.0645		5799.48
ARIMA(3,1,3)	-0.8046	-0.9405	-0.3406	0.9495	0.9650	0.4972	5795.6
ARIMA(2,1,1)	-0.5486	-0.0330		0.6926			5796.38
ARIMA(3,1,2)	-0.5267	-0.5837		0.6766	0.5897	0.1478	5797.7
ARIMA(2,1,3)	-0.4872	-0.0953	0.0458	0.6315	0.0645		5799.48

Table 7.1: Analyzing Valid Models

The Akaike information criterion (AIC) is a approach to find out the relative quality of statistical models for a given set of data.Lower the value of AIC,the batter the model .From the table x we clearly see that ARIMA(2,1,2) is the most suitable model for this particular time series .

Using ARIMA(2,1,2) we forecast next 30 days of share price .Bellow the forecasted time series with standard error .

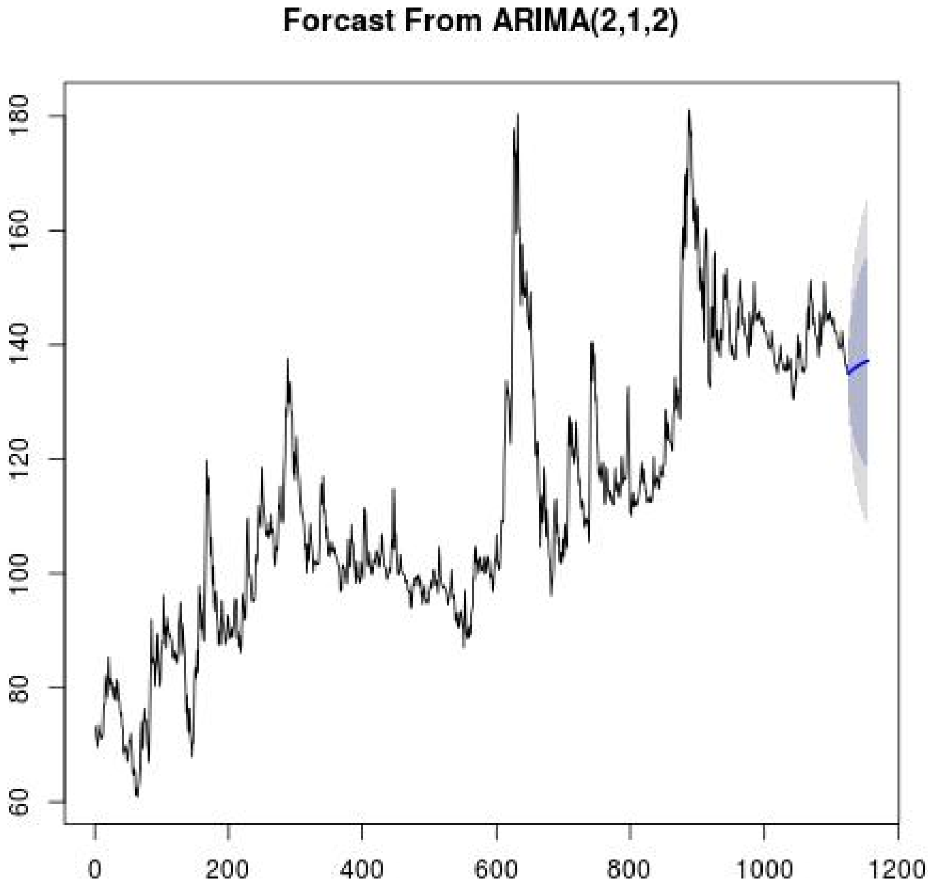

Figure 7.1: Forecast From ARIMA(2,1,2)

Now we will check the accuracy of our predicted ARIMA model by comparing the actual values vs predicted values.

Trade Date	Actual Value	Predicted Value
02-04-2017	137.00	134.94
03-04-2017	137.80	135.16
04-04-2017	137.60	135.21
05-04-2017	140.00	135.36
06-04-2017	138.30	135.44
09-04-2017	137.00	135.56
10-04-2017	136.70	135.65
11-04-2017	136.40	135.75
12-04-2017	135.60	135.84
13-04-2017	135.90	135.93
16-04-2017	135.50	136.01
17-04-2017	136.60	136.01
18-04-2017	137.50	136.18
19-04-2017	136.10	136.26
20-04-2017	138.10	136.33
23-04-2017	135.60	136.41
24-04-2017	135.70	136.48
25-04-2017	136.30	136.55
26-04-2017	136.80	136.61
27-04-2017	137.60	136.68
30-04-2017	134.60	136.74
02-05-2017	133.20	136.80
03-05-2017	130.70	136.85
04-05-2017	130.50	136.91
07-05-2017	132.40	136.96
08-05-2017	133.30	137.02
09-05-2017	135.50	137.07
11-05-2017	134.00	137.12
14-05-2017	138.00	137.16
15-05-2017	141.80	137.21

Table 7.2: Actual Values Vs Forecast Values

The Root Mean Square Error(RMSE) of this series is 2.71 which is poor . We also analysis the other company and we find they really don't show any seasonality like APEX FOODS LIMITED and we also calculate the RMSE for the other company . The results are bellow -

Company Name	RMSE
APEX FOODS LIMITED	2.71
ACI FORMULATIONS LIMITED	7.14
ANWAR GALVANIZING LIMITED	1.15
NATIONAL TEA COMPANY LIMITED	33.27

Table 7.3: RMSE Of Companies

Chapter 8

Future Approch

We r supposed to do our work with multivariate process but we cannot do this for lack of time.We know seasonal effect analysis are available for univariate ARIMA and which call seasonal ARIMA .And there is also a ARIMA model for multivariate time series .But there is no research have done yet to find the seasonal effect on multivariate time series with ARIMA model. We also think, if we continue doing forecast using previous year's data with ARIMA model it will not bear a good solution thus it will be improper.

Chapter 9

Conclusion

This elaboration that all methodical context are based on the ideas that purely probable in near future. It works in short term field works rather than the long ones.But we assure that it helps to solve contemporary complex but not rich enough to provide us actual future queris.

we have use univariate time series in different fields got satisfactory result. There are many such places where the ARIMA model works quite good, but it is not enough effective in DSE.However the 4 or 5 companies where we conduct this model simultaneously by seasonal data, leaves no remarkable mark .

Appendix A

G.C.TAO,Modeling Multiple Time Series with Application,Journal of American Statistical Association,1991.
Abdulkader Aljandali,Multivariate Methods and Forecasting with IBM,2017
Stock Price Forecasting Using Information from Yahoo Finance and Google Trend
Modeling Multiple Time Series with Application
http://people.duke.edu/ rnau/arimrule.htm
https://www.r-bloggers.com/forecasting-stock-returns-using-arima-model/
https://onlinecourses.science.psu.edu/stat510/node/66
http://halweb.uc3m.es/esp/Personal/personas/amalonso/esp/TSAtema6.pdf
http://www.maths.qmul.ac.uk/ bb/TS$_{C}hapter7.pdf$
$http://www.maths.qmul.ac.uk/ bb/TS_{C}hapter7_{2}.pdf$

Bibliography

[1] N. M. David Enkea*, Manfred Grauerb, *Stock Market Prediction with Multiple Regression, Fuzzy Type-2 Clustering and Neural Networks*, 2011.

[2] P. S. Äurka Peter1, *ARIMA vs. ARIMAX â which approach is better to analyze and forecast macroeconomic time series?*, 2012.

[3] S. S. Hiremath, *Preventive and Community dentistry*, 2nd ed. Elsevier India, 6th July 2011.

[4] A. Katchova, *Time Series ARIMA Models*, 2013.

[5] A. O. A. Ayodele A. Adebiyi., *Stock price Prediction Using The ARIMA model*, â2014.

[6] Y. L. Hung Pham, Andrew Chien, *A Framework for Stock Prediction*, December 11, 2009.

[7] K. U. H. D. Swapnil Jadhav1, Saurabh Kakade2, *Indian Share Market Forecasting with ARIMA Model*, 4th ed. International Journal of Advanced Research in Computer and Communication Engineering, November 2015.