

# IMPLEMENTATION OF CORRELATION AND REGRESSION TECHNIQUES FOR EFFICIENT ANALYSIS OF INDIAN STOCK MARKET

Cerene Mariam Abraham<sup>1</sup>, M. Sudheep Elayidom<sup>2</sup>, T. Santhanakrishnan<sup>3</sup>

<sup>1</sup>Research Scholar, Computer Science and Engineering, CUSAT, Cochin, India

<sup>2</sup>Associate Professor, Computer Science and Engineering, CUSAT, Cochin, India

<sup>3</sup>Defence Research & Development Organization, Kerala, India

## Abstract

Stock markets are the most happening places for most of the financial transactions in a country. Derivatives markets, a part of the stock market has become one of the emerging platforms for investors due to the low risk involved. The returns on such investments depend upon your accuracy of prediction of the future value of assets. A number of studies have been done to understand the effect of certain factors on the future value of share prices. Some of the often checked factors are earnings per share, inflation, book value, economic strength, etc. This study revolves around the influence of the less explored but very important factors such as Open interest, Number of Contracts and Deliverable quantity on the share prices. The method of study involved statistical techniques such as Correlation and Regression to study ninety days data of five large cap stocks.

**Keywords:** Open Interest, Deliverable quantity, Regression analysis and Correlation

\*\*\*\*\*

## 1. INTRODUCTION

Indian Stock Markets – Bombay Stock Exchange (BSE) and National Stock Exchange (NSE) are world's 11<sup>th</sup> and 12<sup>th</sup> ranked stock exchanges in the world having market capitalization of approximately \$1.43 trillion each. These are places where all major economic transactions happen and where shares of listed companies are traded by the public. Derivatives market is a category of the financial market where the prices of shares are predicted for future dates[6]. This type of market is considered to have lower risk when compared to other markets as it gives protection against adverse movements in future prices. Hence the future derivatives markets have become increasingly popular and most commonly used in the financial world.

As the derivatives market mainly revolves around prediction of future prices of shares, it is very important to understand the factors influencing the share price and its future value[4]. This paper aims at studying the derivatives market and analyzing the factors influencing the future price of shares through statistical approach.

## 2. METHOD OF STUDY

In a stock market, the investors look at various factors such as earnings per share, inflation, book value, economic strength, etc. to predict the future value of a share[1]. As part of the study, these factors and others were studied. Based on the previous data of price variations of stocks and the knowledge shared by the stock exchange members, the following factors were picked up to study their effect on the future price of shares.

- Open interest

- Number of Contracts
- Deliverable quantity

These factors were understood in detail and explained briefly as follows:

### 2.1 Open Interest

Open interest is the total number of Futures and Options contracts that is open or outstanding for that particular day. Futures and Options are two types of derivatives which are commonly traded. Futures contract is an agreement between two parties to buy or sell a particular share at a specified price for a future date. Options are contracts which give the right but not obligation to buy or sell a particular asset at a certain price on or before a specified date.

### 2.2 Number of Contracts

It is the total number of trade contracts both buy and sell.

### 2.3 Deliverable Quantity

Deliverable volume or Deliverable quantity is the total quantity of shares moving from the demat account of one group of people who are selling to the demat account of another group of people who are purchasing them, on a particular day. Dematerialized or Demat account is opened by every investor when they are registering with an investment broker. Transactions such as buying and selling shares happen through this account during trading.

Statistical analysis using Correlation and Regression techniques were used to understand the dependability among

each factors and study the effect of these factors on the future price of shares. Study was done on the share price variations of large cap stocks. Large cap stocks are shares of large and well-established companies which are having a strong market existence and therefore considered as safe speculation shares. Following are the large cap stocks taken for the study- LT,SBIN,TATA MOTORS, HINDALCO and AXIS BANK.

### 3. STATISTICAL ANALYSIS

#### 3.1 Correlation Analysis

Correlation is a statistical technique used to find the strength of association between two variables. It tells you whether dependency of one variable over the other variable exists or not. Coefficient of correlation ( $r$ ) gives the strength of relation between two variables. It varies from -1 to +1. -1 indicates a strong negative correlation between the two variables, i.e., as variable increases its value, the other decreases its value and vice versa. +1 indicates a strong positive correlation between two variables, i.e., as value of one variable increases, the value of the other variable also increases and vice versa.

R software is a programming language most widely used by statisticians[8]. It is an open source programming language used for statistical computing and graphics. Correlation analysis using R programming was done on ninety days of data on the above mentioned five large cap stocks[5]. Below are the steps followed for analysis.

- The required data were extracted. Each variable was named such as X for Open Interest, Y for number of contracts, Z for Deliverable quantity and P for price of the stock. This was then saved in .txt file.
- The txt file was read in R using command `read.table()`.
- Now the correlation between the variables X,Y,Z and P is checked using the correlation function `cor(variable_1, variable_2, method="method")`.

#### 3.2 Result of Correlation Analysis

The correlation coefficients between V2 and V3, V3 and V4 and V2 and V4 are -0.15, -0.44 and -0.33 respectively. The negative correlation coefficient values show that there exists an inverse relation between these variables, i.e., as one increases, the other decreases.

The plot function is used to graphically plot relationship between these variables and the graph of each company is as follows. X indicates the variable Open Interest, Y indicates the Total number of Contracts, Z indicates the Deliverable quantity and P denotes the Stock price. D1 to D5 shows the selected 5 companies.

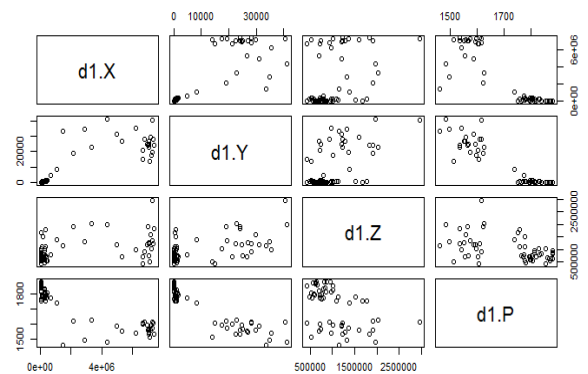


Fig -1: For the company LT.

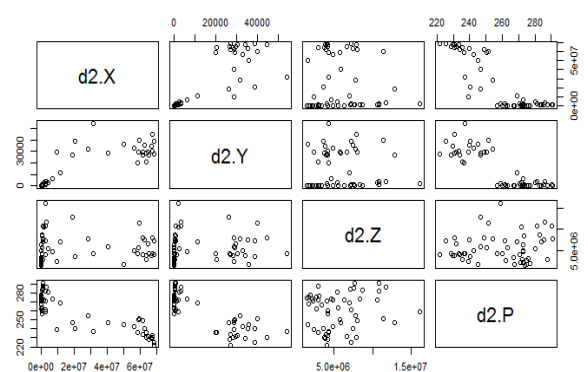


Fig -2: For the company SBIN.

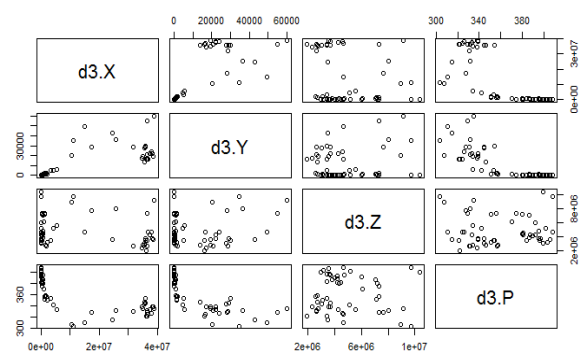


Fig -3: For the company TATAMOTORS.

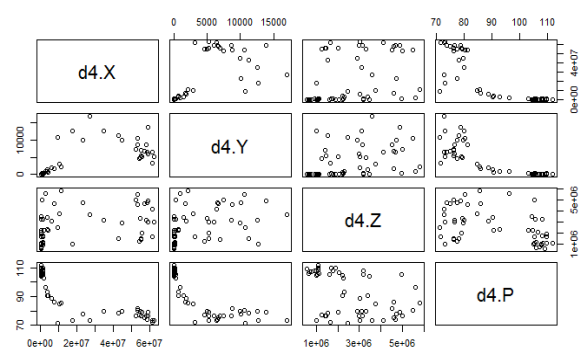


Fig -4: For the company HINDALCO.

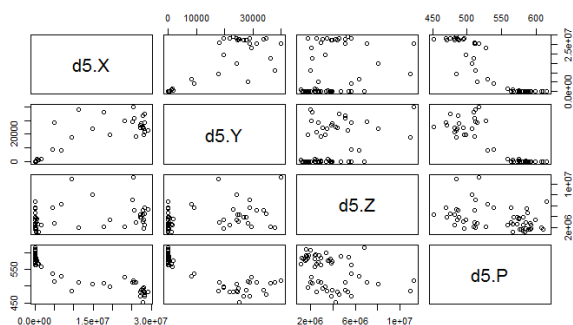


Fig -5: For the company AXISBANK.

### 3.3 Regression Analysis

Regression is a statistical technique used to estimate the relationship between dependent (response) variable and independent (predictor) variables. Based on the past data, it will provide the best fit line or equation connecting the dependent variable and the independent variables[3]. The coefficient of determination,  $R^2$ , pronounced as R squared stipulates how well the data fits the statistical model or equation. An  $R^2$  value of 1 indicates that the regression model best fits the data and an  $R^2$  value of 0 indicates that the regression equation does not fit the data at all. Any value greater than 0.3 indicates a good fit model.

For conducting the regression analysis, 3 months data of the 5 large cap stocks is considered. In this case, the dependent variable is the Price of share and the independent variables are Open interest, number of contracts and Deliverable quantity. The test for significance of regression is carried out using ANOVA (Analysis of Variance). Below are the null and alternate hypotheses –

$$H_0: \beta_1 = \beta_2 = \beta_3 = 0 \text{ and } H_1: \beta_j \neq 0 \text{ for at least one } j.$$

$\beta_j$  are the coefficients of independent variables in the regression equation.

Null hypothesis states that the response variable (Price of share) is not dependent on the predictor variables (factors). Alternate hypothesis states that response variable is dependent on the predictor variables.

Based on the analysis conducted, below are the regression coefficients of the predictor variables for each of the large cap stocks[2]. X denotes Open interest, Y, Number of Contracts and Z, Deliverable quantity. Dependent variable is taken as P, the price of the stock.

Coefficients <sup>a</sup>						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1811.082	15.873		114.101	.000
	Z	7.946E-006	.000	.031	.494	.623
	X	-1.196E-005	.000	-.283	-3.037	.004
	Y	-.007	.001	-.698	-6.735	.000

a. Dependent Variable: P

Fig -6: Regression analysis of LT

Coefficients <sup>a</sup>						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	269.586	2.648		101.810	.000
	X	-3.670E-007	.000	-.521	-4.495	.000
	Y	.000	.000	-.406	-3.498	.001
	Z	8.506E-007	.000	.137	2.259	.028

a. Dependent Variable: P

Fig -7: Regression analysis of SBIN

Coefficients <sup>a</sup>						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	382.663	7.935		48.225	.000
	X	-7.214E-007	.000	-.382	-2.541	.014
	Y	-.001	.000	-.463	-3.113	.003
	Z	-4.566E-007	.000	-.032	-.320	.751

a. Dependent Variable: P

Fig -8: Regression analysis of TATAMOTORS

Coefficients <sup>a</sup>						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	106.361	1.675		63.496	.000
	X	-2.994E-007	.000	-.514	-6.516	.000
	Y	-.001	.000	-.417	-5.213	.000
	Z	-1.205E-006	.000	-.124	-1.988	.052

a. Dependent Variable: P

Fig -9: Regression analysis of HINDALCO

Coefficients <sup>a</sup>						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	585.116	4.803		121.819	.000
	X	-2.372E-006	.000	-.629	-5.557	.000
	Y	-.001	.000	-.298	-2.560	.013
	Z	-1.360E-006	.000	-.064	-1.131	.263

a. Dependent Variable: P

Fig -10: Regression analysis of AXISBANK

### 3.4 Result of Regression Analysis

As seen in the above tables, the p-value of the corresponding regressors is not zero, hence showing that the predictor variables are significantly impacting the response variable. Below are the coefficients of determination for the regression model of each of the companies.

Table -1: Name of the Table

Company	R square	Adjusted R <sup>2</sup>	Standard Error of Estimation
LT	.867	.860	49.18218
SBIN	.813	.802	8.82758
TATAMOTORS	.628	.606	19.35097
HINDALCO	.835	.826	6.11781
AXISBANK	.860	.852	17.82672

As  $R^2$  value is greater than 0.6 for the regression models of each of the companies, all are good fit models and the total response variation is explained by each of the model.

#### 4. DISCUSSIONS

The main findings of the study are as follows-

- There exists an inverse relationship among the variables – Open interest, number of contracts and Deliverable quantity, with the stock price that is, as one increases, the other decreases and vice versa.
- Each of the factors – Open interest, Number of contracts and Deliverable quantity is significantly impacting the Share price.

To arrive at the findings, study was done on ninety days data for 5 large cap stocks. Statistical techniques such as Correlation and Regression analysis were conducted on the data. Hypothesis testing using ANOVA was also done which proved that the factors were significantly impacting the share price.

In a stock market, the investors look at various factors such as earnings per share, inflation, book value, economic strength, etc. to predict the future value of a share. But the above factors are not explored much to see their impact on the share price. The results of the study can be used in Derivatives markets by Investors as one of the guides for buying and selling shares. But like every statistical study, we cannot go by this result blindly as there is 5% probability of failure for this experiment.

#### ACKNOWLEDGEMENT

The author wishes to express her gratitude to Mr. Manoj P. Michel, Cochin Stock Exchange, who shared his profound knowledge of the equity market and for his useful discussions.

#### REFERENCES

- [1]. Ashok kumar, S. Murugan. Performance Analysis of Indian Stock Market Index using Neural Network Time Series Model. IEEE; 2013.
- [2]. Cerene Mariam Abraham, M. Sudheep Elayidom, T. Santhanakrishnan, "Design and Implementation of Efficient Regression Analysis Techniques in Derivative Market", The International Journal Of Science & Technoledge (ISSN 2321 – 919X), Vol. 4, Issue 11, November 2016.
- [3]. Samprit Chatterjee, Ali S. Hadi. Regression Analysis by Example. 5th ed. Wiley Series in Probability and Statistics; 2015.
- [4]. Lu H, Han J and Feng L, "Stock Movement Prediction and N-dimensional Inter Transaction Association Rules", ACM SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery, 1998, pp 12.1-12.7.
- [5]. Cerene Mariam Abraham, M. Sudheep Elayidom, "Implementation of Correlation Techniques for the Equity Market in India", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 4, Special Issue 1, June 2015.

[6]. Snehal Bandivadekar and Saurabh Ghosh, "Derivatives and volatility on Indian Stock Markets," Reserve bank of India Occasional Papers, vol. 24, No. 3, Winter 2003.

[7]. Nirmal Mohanty, "Cost of trading in Stock Exchanges: a Perspective", NSE Working paper, Nov 2011.

[8]. Mark gardener, "Statistics for ecologists using R and excel: Data Collection, exploration analysis and presentation", Pelagic Publishing.

[9]. Pitabas Mohanty and Supriti Mishra, "Run-up in Stock Prior to Merger and acquisitions announcements: Evidences from India," WP/12/2014, NSE, March 2014.