International Journal of Mathematics Trends and Technology
ISSN: 2231 – 5373 /doi:10.14445/22315373/IJMTT-V67I7P515
Volume 67 Issue 7, 118-134, July, 2021

Logistic Regression Model For Predicting Performance of S&P BSE30 Company Using IBM SPSS

Mrinalini Smita*

Assistant Teacher, Mathemaics,
St. Xavier's College, Ranchi , Jharkhand(India)

ABSTRACT:

Stock price prediction, the method of determining future values of a company's stocks and other financial values, is an important topic in finance and economics which has captured the interest of researchers over the years to develop predictive models. A stock performance can, do some extent, be analysed based on financial indicators presented in the company's annual report. Financial ratios help to form the basis of investor stock price expectations and hence, influence investment decision-making. Stock market prediction with the help of binary logistic regression using relation between financial ratios and stock performance has been recognized that financial ratios can enhance an investor's stock price forecasting ability. The purpose of this study is to apply statistical methods to survey and analyze financial data in order to develop a simplified model for interpretation. The main purpose of the study to apply logistic regression model (binary) model for classifying S&P BSE 30 stocks into two categories GOOD or POOR performance stock. The logistic regression model, by applying variable to logistic curves, can be used to predict the likelihood of good performing stocks.

Keywords*: financial ratios, stock performance, stock price prediction, logistic curves ,logistic regression model.*

INTRODUCTION:

The movements of stock prices and stock indices are influenced by many macro-economic variables such as political events, policies of the corporate enterprises, general economic conditions, commodity price index, bank rate, loan rates, foreign exchange rates, investors' expectations, investors' choices and the human psychology of stock market investors. **Miao[1]**Hence to develop predictive models for stock market prediction is a difficult task due to the uncertainty involved in the movement of stock market. That is why it requires continuous improvement in forecasting models..Regression analysis is one of the most useful and the most frequently used statistical methods. Among the different regression models, logistic regression plays a particular role**.Ngunyi [2]** Logistic regression extends the ideas of linear regression to the situations where the dependent variable Y is categorical. Logistic Regression is applied to categorize a bunch of independent variables into either two or more mutually exclusive classes. **Ali[3]** The aim of the regression methods is to describe the relationship between a response variable and one or more explanatory variables. Logistic regression is a type of generalized linear model (GLM) for response variables where regular multiple regression does not work very well. Logistic regression seeks to

- **MODEL** the probability of an event occurring depending on the values of the independent variables , which can be categorical or numerical .

- **ESTIMATE** the probability that an event occurs for a randomly selected observations verses that the probability that the event does not occur.
- **PREDICT** the effect of a series of variables on binary response variables.
- **CLASSIFY** observations by estimating the probability that an observation is in a particular category (such as GOOD or POOR performance of Company of S&P BSE 30 in our case).

The main purpose of the study to apply logistic regression model (binary) model for classifying stocks into two categories GOOD or POOR performance stock. A company's stock is classified as GOOD if its shareholder's profit is high compared to market returns gained from the S&P BSE30 benchmark index. Thus, logistic regression is applied to categorize a bunch of independent variables into either two or more mutually exclusive classes and can be used to predict the likelihood of good performing stocks by applying variable to logistic curves.

❖ OBJECTIVE:

- To make awareness of application of logistic regression model (LRM) using IBM SPSS software in S&P BSE30 in stock market performance prediction based on financial ratios.
- To prove this logistic regression model (LRM) through IBM SPSS as a stepping stone in future prediction technologies.

❖ LITERATURE REVIEW:

Reference[4] suggested that across the United States a number of studies have observed a cross-sectional relationship existing between stock returns and fundamental variables. These fundamental variables are constituted on the basis of analysis of different financial ratios like earnings yield, cash flow yield, book to market value and size of firm and they have proven out to be significant indicators in determining the performance of stocks and have been observed to exercise substantial impact on prediction of stock returns. European based researches have also identified similar findings to have occurred in the European Markets. The model generated is being used extensively throughout the European Markets e.g. UK, France and Germany allowed sensible prediction of returns.

Reference[5]applied MLP technique to successfully predict advertising and marketing trends, macroeconomic data, financial time series forecasting and stock market trends respectively. **Harvey [6]** studied fresh equity markets that have emerged in Europe, Latin America, Asia, the Mideast and Africa and provided an innovative set of opportunities for investors. He elaborated that high expected returns and increased volatility are the features common to these markets. More importantly, he identified significant reductions in unconditional portfolio risk of world investors due to low correlations with developed countries' equity markets. In contrast, he identified that for explaining the cross section of average returns in emerging countries, standard global asset pricing models that undertake complete integration of capital markets are not suitable. His investigation of the predictability of returns revealed that local information deeply influence the emerging market returns.

Studies carried out by **Jung and Boyd [7]** suggested that in predicting stock performance of UK stock prices the models of stock performance tend to be fairly effective. According to Cheng *et al.* (1996); Van and Robert (1997) to create a framework for financial time series, artificial neural networks (ANN) have been fruitfully used.

Al-Loughani [8] while conducting a study on the Kuwait Stock Exchange witnessed the emergence of institutional investors as the leading performers in the market. They estimated that analysis of financial ratios and Logistic Regression techniques play a significant role in the forecasting of out-performing stocks.

Reference[9] forecasted the trends in the returns on market indices of the Taiwan Stock Exchange by applying a probabilistic NN model. The findings were then equated with the generalized methods of moments (GMM) along with the Kalman Filter.

References([10],[11]) also upheld the advantages of Logistic Regression by confirming that through the accumulation of a suitable association function to the standard linear regression model, there may either be continuous or discrete variables, or any combination of both types, which don't essentially have normal distributions.

Reference[12] applied a series of NN models and linear regression models to Istanbul stock exchange and New York Stock Exchange respectively. The composite index data from the periods ranging from 1990-2002 of the Istanbul Stock Market and from 1981-1999 in the New York Stock Exchange were analyzed and successfully predicted the role of trading volume in the respective stock exchanges.

Reference[13] emphasized that Logistic regression can come in handy in conditions where prediction of the existence or deficiency of an outcome or feature is dependent on values of a set predictor variable. In this, the Logistic regression model is very much like the linear regression model, however, it is more convenient where there exists a dichotomy of dependent variable. The models with Logistic regression coefficients have a tendency to use estimated odd ratios for every independent variable. A multi-variant regression can be formed between dependent and independent variables under Logistic regression.

Reference[14] conducted a wide-ranging evaluation of various studies related to bankruptcy prediction problems and discovered that neural network is the highly-accepted mode of statistical modelling for prediction of stock performance.**Ogut[15]**figured out that data-mining methods such as Artificial Neural Networks and Support Vector Machines are more appropriate to spot manipulation of stock price as compared to econometrical methodology's and multivariate data analysis techniques for example Logistic Regression Model and Multiple Discriminate analysis, this is because data-mining techniques perform better and accurate classifications rather than multivariate techniques. These studies proposed different binary classification method, founded on genetic algorithms, for forecasting corporate failure and suggested for authentication an empirical analysis as its prediction strategy.

Other methods for forecasting accuracy were equated by **Min[16]**, like multivariate data analysis such as MANOVA, Factor Analysis, Structural Equation Modelling and Multiple discriminate analysis, logistic regression, decision tree, and artificial neural network, and the results indicated that the binary classification technique which they suggested, has the tendency to prove out to be an encouraging substitute to prevailing techniques for predicting insolvency.

Reference[17] by using the neuro computational model to predict the stock exchange movement in emerging markets, analyzed that quasi Newton training algorithm is effective as compared to training algorithms and exhibit few forecasting errors. He concluded that the results of neuro-computational model are much more reliable than the results of Logistic regression model and Auto regressive integrated moving average. Studies conducted on the Taiwan Stock Exchange (TSEC) by **Chen [18]** explored the reasons of financial distress predication model. In his study, he made use of 37 financial rations on a sample consisting 100 listed firms at the Taiwan Stock Exchange. In order to exclude or combine the variables he carried out PCA (Principal Component Analysis). **Chen [18]** suggested that continuous and robust growth in stock and derivative securities markets throughout the world has allowed for quick market developments and enterprise operating status is to be disclosed periodically on financial statement. However, it is very unfortunate when company executives purposefully design financial statements up, it becomes very difficult to point out any chances of financial distress either in the short or long run.

Reference [19] supported the notion that predicting stock exchange rates is an important financial problem and deserves due attention and recognition. They advocated that in recent times, for generating genuine forecasting results, several neural network models and hybrid models have been introduced that attempted to beat the old linear and nonlinear approaches. Their study assessed the efficacy of neural network models that have the reputation to be dynamic and useful in stock-market predictions. They analyzed the multi-layer perceptron (MLP), dynamic artificial neural network (DAN2) and the hybrid neural networks models to objectify the autoregressive conditional hetero-scedasticity (GARCH) for extracting different input variables. Each model contained two major points of discussion: Mean Square Error (MSE) and Mean Absolute Deviate (MAD) using real exchange daily rate values of NASDAQ Stock Exchange index.

References([20],[21]) investigated that logistic regression (LR) when applied on different financial ratios as various independent variables specify significant mark on the performance of stocks which are being actively exchanged on the Indian stock exchange. This research utilized a sample of ratios from thirty big market capitalizing companies over duration of four years. The research conducted an investigation on 8 financial ratios, which categorized these companies in to two groups – “good” or “poor” –, up to an accuracy of 74.6%, gauging their market rate of return. The study declares that the framework created can improve the stock price forecasting aptitude of the investors. However, other Macro-economic variables that can exert significantly affect the stock prices, were not considered. The study elucidated the applied inferences of using the Logistic Regression model in order to forecast the prospects of good performing stocks. Author’s claim that to expand the capability for selecting good stocks, this model is useful for investors, fund managers, and investment.

Reference [22] also applied multi nominal logistic regression to forecast stock performance in Indian market and found similar results.

Reference [23] studied that to provide a preliminary guideline for short term investors the early forecasting of the direction of share market may become imperative as warning strategy to ultimately prevent financial distress for long term shareholders. Most of the stock prediction researches emphasize on using macroeconomic indicators, such as CPI and GDP, to gauge the prediction model.

Thus, Existing literature indicates that logistic regression (LR) has been rarely used to build a model for predicting out-performing shares. Logistic regression has been used mostly for predicting financial distress and business failure. It has not been used for predicting share performance in India. In terms of investment destination in share, India is a top performing emerging market. In this context, the present study will provide useful information to shareholders and potential investors to enable them to make good decisions regarding investments.

❖ RESEARCH METHODOLOGY:

1. TYPE OF REASEARCH

This study is data based ***empirical or experimental*** research coming up with conclusions which are capable of being verified by observations or experiments.

2. TOOLS (Statistical tools) USED IN THE STUDY:

To forecast the movements of S&P BSE Sensex, financial ratios-

- Earnings Per Share(EPS),
- Price Earnings ratio(PE),
- Price Earning Growth(PEG),
- Earnings Before Interest Tax Depreciation And Amortization(EV/EBITDA),
- Price to Book Value(P/B) and
- Price to Sales

are considered and classification prediction was done using IBM SPSS27.0

(***Statistical Packages for Product and Services Solutions***) .

3. SAMPLE SIZE (Sampling Design) OF THE STUDY:

Based on financial ratios , as on 1st March ,2019 to 1st March,2020, S&P BSE30 Sensex are considered since these signify the POOR / GOOD performance of stock market movements of S&P BSE30 Sensex in India.

Sources of Data:

The most of the secondary data of the financial ratios of S&P BSE Sensex of the financial year 2019-2020 were collected from the websites of BSE.

Period of the Study:

The study focus on the behaviour of stock price / index returns of one year from 1st March 2019 to 1st March 2020, just before the COVID -19 crisis.

❖ SIGNIFICANCE OF THE STUDY:

- This study will educate laymen for proper investment with substantial risk in stock market.
- We want to make investment as 'social revolution ' by spreading awareness of capital investment in stock market.
- It will help in boosting the confidence of stakeholders in financial industry to do more business with less risk.
- It will enable public to invest wisely in stock market.

❖ LOGISTIC REGRESSION MODEL:

In Logistic Regression (LR), instead of predicting ***the value of the variable Y*** from a predictor variable X or several predictor variable Xs, we predict the ***Probability of Y occurring*** , given known values of Xs. Hence the logistic regression model is a type of generalized linear model (GLM) that extends the linear regression model by linking the range of real numbers to 0-1 range. Whereas OLS (ordinary Least Square) regression uses ***normal probability Theory,*** Logistic Regression uses ***Binomial Probability theory.*** In Logistic regression, instead of predicting ***the value of the variable Y*** from a predictor variable X or several predictor variable X's , we predict the ***Probability of Y occurring*** , given known values of X's. This statistical method is not used to model regression problem. It is a supervised leaning used to solve classification. The term "***logistic"*** refers to the "***log odds***" (the probability that is modelled). Logistic regression model the relationship between the dichotomous dependent variables depending upon odds ratios ODDS RATIO (the ratio of probability of happening to the probability of non-happening) . It means odds increases, probability increase and vice versa. Probability ranges from 0 to 1 where as odds ranges from 0 to $+\infty$. In logistic regression, the dependent variable is a *log odds or logit*, which is the natural log of odds. For a variable in logistic regression, it represents ***how the odds change with one unit increase in that variable holding all others variables constant.***

Logistic regression model the relationship between the dichotomous dependent variables. This model depend upon odds ratios. ODDS RATIO for a variable in logistic regression represents ***how the odds change with one unit increase in that variable holding all others variables constant.***

$$\frac{p(x)}{p(1\text{-}x)} = [\exp(-X^T\beta)]^{-1} = \mathbf{ODDS}$$

Taking natural log both sides

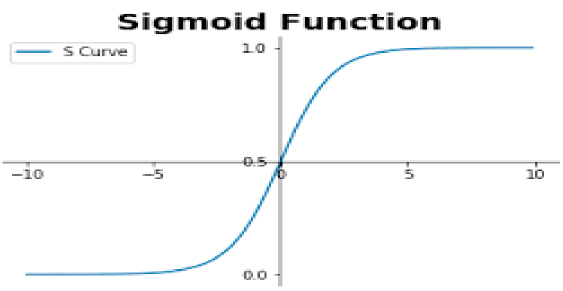
$$\log\frac{p(x)}{p(1\text{-}x)} = [-(X^T\beta)]^{-1}$$

Above transformation in log is known as ***logistic transformation***

Let the condition of probability be $\Pr((Y = 1|X) = p(X)$

Here **1** is not the number but ***a class or category*** so we need to model a probability using a curve where the predictor domain X can be anything and the range of $p(X)$ (or a condition of probability that y should given x) is between 0 and 1. Classification in case of linear regression is not possible as h(x) always gives real values . So we can apply another function on the linear function so that we can use the result for classification.

That another function is ***Logistic Function(Sigmoid Function)***



In the graph , as $z \to \infty, g(z) \to 1$

and $z \to -\infty, g(z) \to 0$

Hence 1 and 0 are the upper and lower bound respectively for the sigmoid function, 0.5 is the value that the sigmoid function takes when x =0.

We say that *sigmoid function* has ***squashing effect*** which means that for any value that x takes, the value of sigmoid function taken is always between 0 and 1.

$$-\infty < x < +\infty \Longrightarrow 0 < \sigma(x) < 1$$

- **FORMULA OF SIGMOID FUNCTION:**

$$\boldsymbol{g}(\boldsymbol{z}) = \boldsymbol{\sigma}(\boldsymbol{z}) = \frac{\mathbf{1}}{\mathbf{1}+\boldsymbol{e}^{-\boldsymbol{z}}}$$

Just like in regression h(x), in logistic regression for classification, we have

$$h(x) = g \sum \beta_i x_i)$$

$$= g(\beta^T X) \quad \text{[\textbf{\textit{Matrix notation}}]}$$

Or $\qquad \sigma(z) = \frac{1}{1+e^{-\beta^T x}}$

We can use a linear function of β, pass it through the ***Sigmoid function*** (S-shaped curve) and use it for ***classification.***

- ***<u>Properties of sigmoid function</u>***:

- As $g(z) \rightarrow 1$ as z$\rightarrow \infty$ and $g(z) \rightarrow 0$ z$\rightarrow -\infty$,
- Derivative of g(z) $= \frac{1}{1+e^{-z}}$ is given by

$$g'(z) = \frac{1}{(1+e^{-z})^2} \cdot e^{-z}$$

After simplification,

$$g'(z) = \frac{1}{1+e^{-z}}\left(1 - \frac{1}{1+e^{-z}}\right)$$

$$g'(z) \quad = g(z)(1 - g(z))$$

This ***derivative*** of Sigmoid function is the most attractive feature of Sigmoid function which is extremely simple to compute and making use of it for classification problem.

❖ **LOGISTIC PROBABILITY FUNCTION**:

We have

$$p = \frac{e^{(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots\ldots\ldots+\beta_n x_n)}}{e^{(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots\ldots\ldots+\beta_n x_n)} + 1}$$

$x_1, x_2, x_3, \ldots\ldots\ldots, x_n$ with n explanatory variables satisfying the two conditions

1. The logistic probability function must always be positive (since p>=0)
2. The logistic probability function must be less than 1 (since p<=1)

The logistic function is NON LINEAR. By performing some algebra we can rewrite the logistic expression to obtain the ***Logistic Regression Model.***

PARAMETER ESTIMATION:

The goal of learning is basically to estimate the parameters in order to make predictions. The parameters in the equation of a 2-class classification in logistic regression is a $\beta^{\wedge}$(β-hat) vector as in the log-odds equation. Just as in regression ,we can find a best fitting line using ***Ordinary Least Square*** (OLS) which minimized the squared residuals or errors, in logistic Regression ,we use calculus based function called ***Maximum Likelihood Function.*** It finds the function that will ***maximize our ability to predict the probability of Y based on what we know about x.***

$$l(\beta) = \sum_{i=1}^{n} y_i \ \beta x_i - \log\left(1 + exp^{\beta x_i}\right)$$

This is the final form of ***likelihood function which is to be optimized.*** The goal is to find the value of β that maximizes this function. The solution must be numerically estimated using an iterative process. Perhaps the most popular method for solving systems of nonlinear equations is Newton's method, also called the ***Newton-Raphson method.***

$\beta^{t+1} = \beta^{t} - \frac{\nabla_{\beta}\, l(\beta^{t})}{\nabla_{\beta\beta} l(\beta^{t})}$ [***Newton raphson Equation***]

We need to compute this for t iterations then data will eventually converge to the approximate coefficient vector.

$\nabla_{\beta} l = \sum_{i=1}^{n} [y_i - [p(x_i)].\, x_i$ [***GradientVector***]

This is the ***numerator term*** of Newton- Raphson Equation.

$\nabla_{\beta\beta} = -\sum_{i=1}^{n} p(x_i)(1 - p(x_i)\, {x_i}^{T}\, x_i$ [***Hessian Matrix***]

This is the ***denominator term*** of Newton- Raphson Equation.

Now converting these into Matrix Notation:

$\nabla_{\beta} l = X^{T}\, (\, Y - Y^{\wedge}\,)$

$$\nabla_{\beta\beta} l = -X^{T} P(1 - P)X$$

$\nabla_{\beta\beta} l = -X^{T} W X$ [replacing P(1-P) by W as diagonal matrix]

Substituting these values in Newton Raphson Equation, we have

$$\beta^{(t+1)} = \beta^{(t)} + (X^{T} W^{(t)} X)^{-1}\, X(Y - Y^{\wedge(t)}\,)$$

Then we have to execute it for number of iterations 't' until the value of converges. Once the coefficients have been estimated, we can substitute the values of some feature vectors X to estimate the probabilities of it belonging to a specific class (by choosing a parameter above which it is ***class 1 (GOOD***) and below which it is ***class 0 (POOR***).

SAMPLING FRAMEWORK:

In this study we started with companies of S&P BSE 30 on the basis of market return for every year from 2019-20. Then we analysis the association between stock performance and financial ratios of S&P BSE30 using binomial logistic regression. The key purpose of this study is to apply the logistic regression framework utilizing financial ratios of the listed companies in order to forecast performance in the S&P BSE30 Sensex.

This study, therefore, answer two questions-

1. Can the return of stocks be explained with the help of different financial ratios?
2. Can we analyze stock return using a logistic regression model?

The study also examines the efficiency of ratio as prediction of stock returns.

TABLE 1

A. DEPENDENT VARIABLE

Type of Company (Based on stock market return)	
GOOD	Return above Market return i.e. BSE
POOR	Return below Market return i.e. BSE

B. INDEPENDENT VARIABLES

NAME OF VARIABLE	**DESCRIPTION OF THE VARIABLE**
EPS	EARNINGPER SHARE
PE	PRICE EARNING RATIO
PEG	PRICE EARNING GROWTH
ROE	RETURN ON EQUITY
EV/EBITDA	EARNING BEFORE INTEREST TAX DEPRECIATION AND AMORTIZATION

LOGISTIC REGRESSION RESULTS (OUTPUT IN SPSS):

The result of the logistic regression model using IBM SPSS is given in below tables.

TABLE -2

(Logistic Regression Results)

VARIABLES IN THE EQUATION

(Using SPSS)

	B	S.E	Wald	df	Sig	Exp (B)
Step 1 EPS	**-0.009**	**0.011**	**0.601**	**1**	**0.438**	**0.991**
PE	**-0.116**	**0.095**	**1.482**	**1**	**0.223**	**0.890**
PEG	**0.374**	**0.237**	**2.494**	**1**	**0.114**	**1.453**
ROE	**-0.160**	**0.098**	**2.673**	**1**	**0.102**	**0.852**
EV/EBITDA	**-0.171**	**0.202**	**0.721**	**1**	**0.396**	**0.843**
CR	**2.196**	**0.999**	**4.829**	**1**	**0.028**	**8.988**
BV	**0.248**	**0.199**	**1.558**	**1**	**0.212**	**1.282**
PR. To Sales	**0.690**	**0.651**	**1.112**	**1**	**0.290**	**1.993**
CONSTANT	**0.782**	**2.399**	**0.106**	**1**	**0.744**	**2.186**

a. Variable(s) entered on step 1: Earnings Per Share, Price Earning Ratio ,Price Earning Growth Ratio ,Return On Equity ,Earnings

TABLE- 3

HOSMER AND LEMESHOW TEST

(Using SPSS)

Step 1	Chi –Square	df	Sig.
1	7.040	8	0.532

TABLE-4

CLASSIFICATION ACCURACY

STOCK PERFORMANCE CLASSIFICATION

Observed Predicted

Step 1		POOR	GOOD	%Correct
PERFORMANCE OF A COMPANY	POOR	14	1	93.3
	GOOD	2	13	86.7
Overall Percentage				90.0

The cut value is 0.500

TABLE- 5

MODEL SUMMARY

(Using SPSS)

Step	-2 log likelihood	Cox and Snell R Square	Negelkerke R Square
1	20.396	0.507	0.675

Estimation terminated at iteration number 8 because parameter estimates changed by less than 0.005

TABLE-6

OMNIBUS TEST OF MODEL COEFFICIENTS

(Using SPSS)

Chi Square	df	Sig.
21.193	8	0.007
21.193	8	0.007
21.193	8	0.007

MAIN FINDINGS:

OBSERVATIONS OF TABLES OF LOGISTIC REGRESSION OUTPUTS IN SPSS:

The **table 1** contains the description of dependent and independent variables (eight financial ratios) of the Logistic Regression Model for predicting trends in S&P BSE 30 Stock Performance of the company.

The **table 2 (VARIABLES IN THE EQUATION)** is the most important of all outputs for our logistic regression analysis. It is the **'*heart of accuracy*'** of our question "***Performance of a company (BSE30)-GOOD or POOR"*** . This table shows the contribution of each independent exact variable to the model and its statistical significance. Exp(B) column (the odds ratio) tells us that company of good performance is achieved 90% of the classification. It shows the regression function

Z=0.782-0.009*EPS-0.116*PE+0.374*PEG-0.160*ROE-0.171*EV/EBITDA+2.196*CR+0.248*PBV+0.690*PRICE TO SALES

The table also indicates the test of significance for each of the coefficients in the Logistic Regression Model. However SPSS gives the significance levels of each co-efficient. As we can see , ***PEG(0.374), CR(2.196), BV (0.248)and Price to Sales (P/Sales) are significant , all other variables (EPS(-0.009), PE(0.116),ROE (-0.160)and EV/EBITDA(-0.171)*** are not.

The **Table 3** (**Hosmer and Lemeshow Test)** is similar to Chi- Square Test means model is significant and variables are fitting well to explain the model of the independent variable.

The **table 4** (**CLASSIFICATION ACCURACY)** contains the classification results with almost **90.0%** correct classification of the model for the performance of the company – POOR or GOOD is perfectly good. This table shows the comparison of the observed and predicted performance of the companies and degree of their prediction accuracy. It also shows the degree of success of the classification for this sample.

The subscript '***The cut value is 0.500'*** mean that if the probability of case being classified into 'GOOD' catagory is greater than .500 than that particular case is classified into the 'GOOD' category otherwise the case is classified into the 'POOR' category. The number and percentage of cases correctly classified and misclassified (2 cases) are displayed. It is clear from this table that the poor companies have a **93.3%** correct classification rate whereas good companies have a **86.7%** correct classification rate. Overall correct classification was observed in **90.0%** original grouped cases.

The **table 5** (**Model Summary**) tells us "***how much variation in the dependent variable can be explained by the model".*** This table contains ***Cox and Snell R Square and Negelkerke R square*** which are both methods of calculating the explained variation. These values are sometimes referred to as ***Pseudo R-Square*** .The explained variation in the dependent variable based on our logistic model ranges from **50.7% or 67.5%.**

The **table 6** (**Omnibus Test of Model co-efficients**) is used to check that the new model(with explanatory variables included) is an improvement over base (null) model. It uses Chi-Square test to see if there is a significant difference between the log-likelihood (specifically -2LLs) of the baseline model (41.589)and the new model(20.396). The new model has a significant reduction in -2LL (41.589- 20.396=**21.193**) compared to the baseline model which suggests that the new model is explaining more of the variables in the outcome and is improvement in the model. In this study, we have added all the eight variables , namely , EPS, PE, PEG , ROE , EV/EBITDA, CR , PBV and PRICE TO SALES in one block , therefore , have only one step.

The Table 7 below represents the *predicted performance of S&P BSE30 Company.* The number and percentage of cases correctly classified and misclassified (2 cases) are displayed. The companies with GOOD performance are shown in GREEN while Companies with POOR performance is shown in RED.

TABLE-7

PREDICTED PERFORMANCE OF S&P BSE30 COMPANY(Using SPSS)

TCS	RELIANCE	INFOSYS	HUL	KOTAK MAHINDRA	SBI	ICICI	NESTLE	BHEL	HDFC HOUSING DEV.
0.54	0.06	0.91	0..99	0.98	0.98	0.81	0.99	0.80	0.21

INDUSIND BANK	MAHINDRA &MAHINDRA	ONGC	TATA STEEL	ASIAN PAINT	HERO MOTOR	HINDUSTAN UNILEVER	LARSEN &TURBO	POWER & GRID	SUN PHARMA
0.98	0.75	1.00	0.15	0.25	0.20	0.15	0.24	0.21	0.71

TECH MAHINDRA	TITAN	ULTRA TECH	MARUTI SUZUKI	NTPC	BHARTI AIRTEL	AXIS BANK LTD.	HDFC BANK	ITC	BAJAJ F. LTD.
0.01	0.38	0.17	0.00	0.28	0.00	0.26	0.86	0.98	0.25

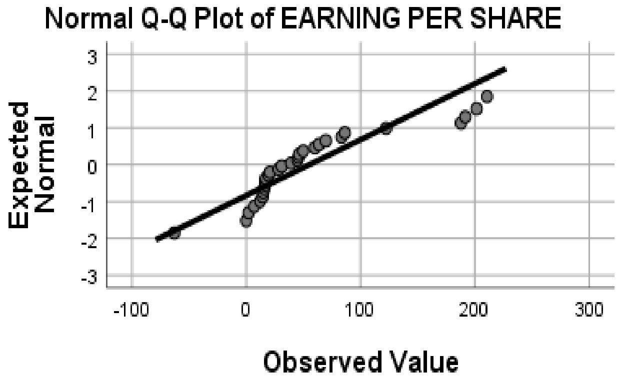
PERFORMANCE OF A COMPANY:

POOR------ GOOD……….

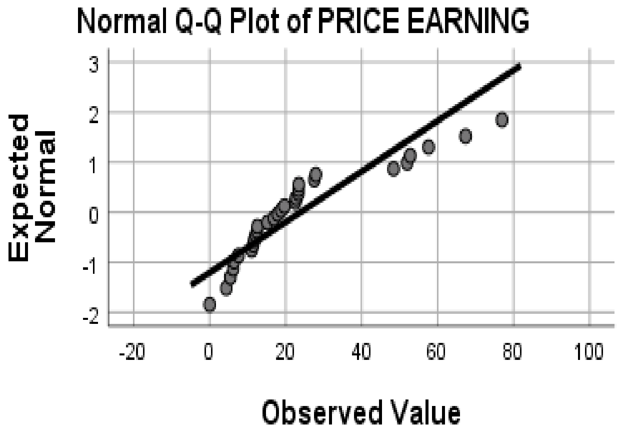
```
                                    Step number: 1
                      Observed Groups and Predicted Probabilities
      4 +                                                                                                    +
        I                                                                                                    I
        I                                                                                                    I
F       I                                                                                                    I
R     3 +                                                                                                    +
E       I                                                                                                    I
Q       I                                                                                                    I
U       I                                                                                                    I
E     2 +                                          g                                                         +
N       I                                          g                                                         I
C       I                                          g                                                         I
Y       I                                          g                                                         I
      1 + p    p                             p   p g                     g   p                             gg+
        I p    p                             p   p g                     g   p                             ggI
        I p    p                             p   p g                     g   p                             ggI
        I p    p                             p   p g                     g   p                             ggI
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        .1        .2        .3        .4        .5        .6        .7        .8        .9        1
  Group:  pppppppppppppppppppppppppppppppppppppppppppppppppppggggggggggggggggggggggggggggggggggggggggggggggggg

                        Predicted Probability is of Membership for good
                                 The Cut Value is .50
                               Symbols:p - poor ; g -good
                             Each Symbol represents .25 Cases
```

Classification Plot in SPSS

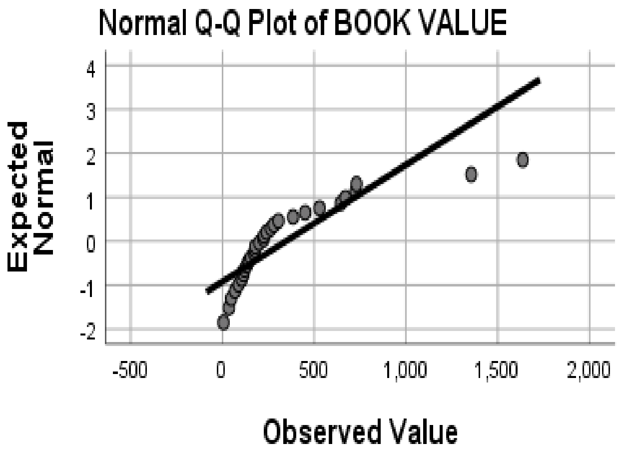
QQ PLOT IN SPSS:



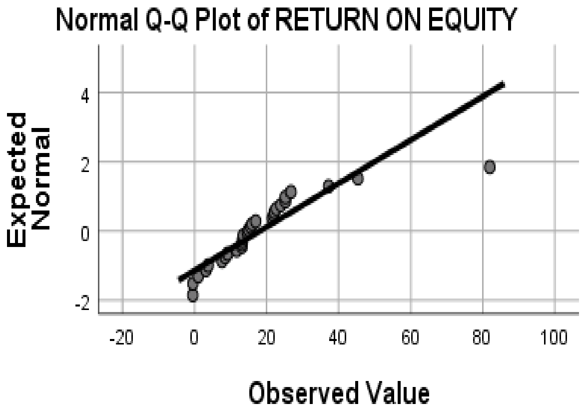
Q-Q Plot for ***Earning per share:***



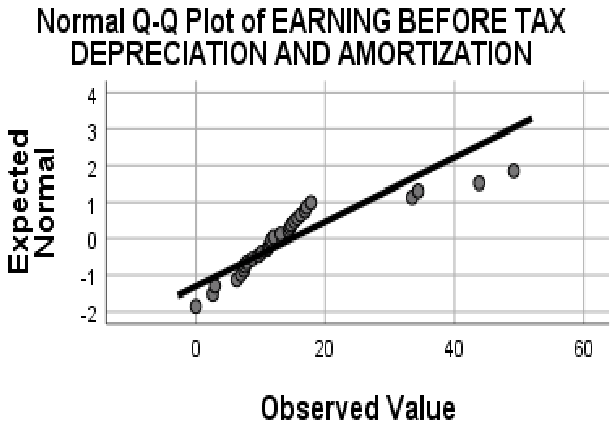
Q-Q Plot for ***Price Earning Growth:***



Q-Q Plot for ***Book Value:***



Q-Q Plot for ***Return on Equity:***



Q-Q Plot for ***Earning before Tax Depreciation and Amortization:***

SUMMARY OF THE RESULTS (FINDINGS):

The result of Stock performance classification shows the overall performance prediction. The performance of the model is analyzed with the help of cross tabulating observed response categories versus predicted performance categories.

In the above table 5 if predicated value is greater than cut off rate it is considered as one or below the cut off rate it is considered as zero. The cut off value is set in this study is equal to 0.5. The analysis of the observed and predicted performance of the companies shows that accuracy of our logistic regression model. In this table, we specify the degree of correct prediction of our model. It shows that ***our model is 90.0% accurately predicting the performance classification.*** The results provided the evidence that earnings per share(EPS), price earnings ratio (PE), price earnings growth(PEG), return on equity(ROE), earnings before interest tax depreciation and amortization (EV/EBITDA), current ratio (CR), price to book value (PBV), price to sales are used as identifier of the company's probability of performing good or poor. Our result predicts ***93.3 percent poor*** performing companies and ***86.7 percent good*** performing companies result accurately. The result of our study shows that ***90.0 percent*** level of accuracy of our model.

CONCLUSION

In this study, the relation between financial ratios and stock performance of the firms has been analyzed with the help of binary logistic regression. The practical implications of using the Logistic Regression method to predict the probability of good stock performance has been shown in the study. The number and percentage of cases correctly classified and misclassified (2 cases) are displayed. It is clear from this table that the poor companies have a **93.3%** correct classification rate whereas good companies have a **86.7%** correct classification rate. Overall correct classification was observed in **90.0%** original grouped cases.

REFERENCES

[1] Miao, K., Chen, F. & Zhao, Z.G., Stock price forecast based on bacterial colony RBF neural network., Journal of Qingdao University (Natural Science Edition), ISSN: 1006-1037 2(11) (2007).

[2] Ngunyi, A., Mwita, P.N., Odhiambo,R.O., On the estimation of properties of logistic regression Parameters, IOSR Journals of Mathematics.,e-ISSN:2278-5728,10(4) (2014) 57-68

[3] Ali, S.S. , Mubeen, M., Lal, I., Hussain , A., Prediction of stock performance by using logistic regression model: evidence from Pakistan Stock Exchange (PSX), Asian Journal of Empirical Research, 8(7) (2018) 247-258 ISSN (P): 2306-983X, ISSN (E): 2224-4425 Journal of Business Intelligence and Data Mining, ISSN:1743-8195,9(2) 145-160.

[4] Fama, E. F., & French, K. R., The cross-section of expected stock returns". The Journal of Finance, 47(2) (1992) ISSN: 0022-1082, 427-465.

[5] Aminian, F., Suarez, E. D., Aminian, M., & Walz, D. T., Forecasting economic data with neural networks". Computational Economics, 28(1) (2006) ISSN: 0927-7099, 71-88.

[6] Harvey, C. R., Predictable risk and returns in emerging markets". The Review of Financial Studies, ISSN: 0893-9454, 8 (1995) 773-816.

[7] Jung, C., & Boyd, R., Forecasting UK stock prices., Applied Financial Economics, ISSN: 09603107, 14664305, 6 (1996) 279-286.

[8] Al-Loughani, N., & Chappell, D., Modelling the day-of-the-week effect in the Kuwait Stock Exchange: a nonlinear GARCH representation., Applied Financial Economics, ISSN: 9603107, 14664305,11(4) (2001) 353-359.

[9] Chen, A. S., Leung, M. T., & Daouk, H., Application of neural networks to an emerging financial market: Forecasting and trading the Taiwan Stock Index., Computers and Operations Research, ISSN: 0305-0548,30(6) (2003) 901-923

[10] Lee, S., Application of likelihood ratio and logistic regression models to landslide susceptibility mapping using GIS., Environmental Management, ISSN: 1095-8630,34(2) (2004) 223-232.

[11] Pardo, J. A., Pardo, L., & Pardo., Minimum Ɵ-divergence estimator in logistic regression models., Statistical Papers, ISSN: 09325026, 16139798, 47 (2005) 91-108.

[12] Yumlu, S., Gürgen, F. S., & Okay, N., A comparison of global, recurrent and smoothed-piecewise neural models for Istanbul stock exchange (ISE) prediction., Pattern Recognition Letters, 26(13) (2005) ISSN: 0167-8655,2093-2103.

[13] Lee, S., Ryu, J., & Kim, L., Landslide susceptibility analysis and its verification using likelihood ratio, logistic regression, and artificial neural network models". Case study of Youngin, Korea, Landslides, CFKO200430710677778, 4 (2007) 327-338.

[14] Kumar, P. R., & Ravi, V., Bankruptcy prediction in banks and firms via statistical and intelligent techniques - A review., European Journal of Operational Research, ISSN: 0377-2217, 180(1) (2007) 1-28

[15] Öğüt, H., Doganay, M. M., & Aktaş, R., Detecting stock-price manipulation in an emerging market: The case of Turkey., Expert Systems with Applications 36(9) (2009) ISSN: 0957-4174, 11944-11949.

[16] Min, J. H., & Jeong, C., A binary classification method for bankruptcy prediction., Expert Systems with Applications, ISSN: 0957-4174,36(3) (2009) 5256-5263.

[17] Mostafa, M. M. (2010). "Forecasting stock exchange movements using neural networks: Empirical evidence from Kuwait". Expert Systems with Application, ISSN: 0957-4174, 37(9), 6302-6309.

[18] Chen, Mu-Yen. (2011) "Predicting corporate financial distress based on integration of decision tree classification and logistic regression", Expert Systems with Applications, ISSN: 0957-4174, 38(9), 11261-11272.

[19] Guresen, Erkam, et al. (2011). "Using artificial neural network models in stock market index prediction", Expert Systems with Applications, ISSN: 0957-4174, 38 (8), 10389-10397.

[20] Dutta, A., et al. (2008). "Classification and Prediction of Stock Performance using Logistic Regression: An Empirical Examination from Indian Stock Market", Redefining Business Horizons: McMillan Advanced Research Series, ISSN: 0007-6813, 46-62.

[21] Dutta, A., Bandopadhyay, G., & Sengupta, S. (2015)." Prediction of stock performance in Indian stock market using logistic regression". International Journal of Business and Information, ISSN: 1728-8673, 7(1), 105-136

[22] Upadhyay, A., Bandyopadhyay, G., & Dutta, A. (2012). "Forecasting stock performance in indian market using multinomial logistic regression". Journal of Business Studies Quarterly, ISSN: 21521034, 3(3), 16-39.

[23] Wang, Y. (2014). "Stock price direction prediction by directly using prices data: an empirical study on the KOSPI and HIS". International Journal of Business Intelligence and Data Mining, ISSN: 1743-8195,9(2) 145-160.