International Journal of Applied Business and Economic Research

ISSN : 0972-7302

available at http: www.serialsjournal.com

Volume 15 • Number 21 • 2017

A Comprehensive LR Model for Predicting Bank's Stock Performance in Indian Stock Market

Nataraja N.S[1], Nagaraja Rao Chilale[2] and Ganesh L[3]

[1]*Alliance University, Bangalore, India*
[2]*Vijaya College, Bangalore University, Bangalore, India*
[3]*Associate professor, Christ University, Bangalore, India*

ABSTRACT

The study focusses on developing a Logistic Regression model to distinguish between "Good" and "Poor" Performance of Bank-stocks which are traded in Indian stock market with regard to the financial ratios. The study- sample comprises of financialratios of 40 nationalised and private banks, for a period of six years. The study ascertains and scrutinizes eleven financial ratios that can categorize the Banksbroadly into two categories as "good" or "poor", up to the accuracy level of 78 percent, based on their rate of return. First, the study predicts the performance of banks by using financial ratios and tries to build the goodness of fit by using Logistic Regression approach. The study also emphasizes that this model can enrich an investor's ability to forecast the price of various stocks. However, the paper confers the real-world implications of Logistic Regression model to envisage the performance of Banks in the stock market. The study reveals that the model could be useful to potential investors, fund managers, and investment companies to improve their strategies and to select the 'out-performing' Bank-stocks.

Keywords: Stock performance, logistic regression, market rate of return, key financial ratios, NIFTY, SENSEX.

1. INTRODUCTION

The Financial stability of a country can be significantly determined by studying the Performance of the banking sector in stock market. In addition to this, involvement of high technology and globalization has made stock market, a highly competitive. All organizations regardless of type of business are affected by this involvement of high technology and globalization. Investors should have virtuous knowledge about

stock exchanges and its role in the stock market. There is high competition not only in domestic banks but also in foreign banks. It is advisable for shareholders and potential investors to procure and apply relevant financial information so as to facilitate them to make good investment decisions in the stock market related to banking sector. Forecasting the performance of various Banks stock is certainly very complicated and difficult. In the history of stock performance literature, no comprehensive, accurate model has been suggested to date for predicting stock market performance. However, to some extent, a stock's performance can be analysed based on financial indicators available from the Bank's annual reports. Financial ratios are significant gears for gauging forthcoming stock performance. Ratio analysis has proven to be one of the key parameters to ascertain the intrinsic value of Banks stock shares (Arun Upadhyay *et. al.,* 2012). Financial ratios are regarded as the baseline for investor's stock price anticipations and, hence, significantly impact investment decisions. The financial ratios used for prediction may vary from one industry to another and at the same time, it may vary from one country to another. Hence, in order to make an accurate decision, decision makers are needed in this industry. The decision makers can use statistical tools for making accurate predictions and face challenges ahead. Thus, selection of appropriate ratios is very decisive in snowballing the success rate of prediction. In this regard, it is important to study the performance of bank - stocks. The focal aim of the authors is to understand and analyse the financial data of various banks and to develop a simplified model for testing the bank's stock performance.

2. REVIEW OF LITERATURE

Many researchers have analysed about the bank performance and found that internal factors relates to banks characteristics and external factors related to economic and legal environment (Athanasoglou, Brissimis & Delis, 2008). In India, Indian stock market has garnered widespread attention from investors worldwide due to its prospective potential and accelerating growth for investment. On account of this growth it has been able to be the cynosure of attention from various quarters. A number of research papers predict stock performance as well as pricing of the stock index across the globe. Avijan Dutta *et. al.,* (June 2012) have used the Logistic Regression model to find the performance of stocks based on few important ratios and observed that the model can enhance an investor's stock price forecasting ability. Ali Ghezelbash and Farshid Keynia (2014) tested the neural network-GARCH model for predicting the combined volatility in the stock index at Latin America, Brazil, Chile and Mexicousing the Artificial Neural network (ANN) model and observed the prediction of performance of GARCH models improved. Also, the studies have shown that the results are robust to different specifications and different measures of volatility consistent with ANN. Performance of a neural network model using the historical prices and macroeconomic data was examined by Wright, Galler and Kryzanowski (2009) to discriminate between stocks with higher performance and those with lower. Results inferred that the neural network correctly classified 72% of the variation of returns, is considered good. Mu-Yen Chen (2011) used the decision tree (DT) classification methods (C5.0, CART, and CHAID) and logistic regression (LR) techniques to implement the financial distress prediction model. He observed DT classification approach obtains better prediction accuracy than the LR approach in short run (fewer one year). On the contrary, the LR approach gets better prediction accuracy in long run (above one and half year). Richard P. Hauser and David Booth (2011) uses three-fold cross validation scheme to compare the classification and prediction of bankrupt firms by robust logistic regression with the Bianco and Yohai (BY) estimator versus maximum likelihood (ML) logistic regression. He proved that robust logistic regression method can significantly improve the classification and prediction of bankrupt firms. Sittichai

Puagwatanaa and Kennedy D Gunawardana(2005) developed a model to predict business failure in Thailand particular in technology industry. The model was developed by using the stepwise logistic regression and using financial information of private limited companies. They concluded that financial ratios are useful analytical techniques for forecasting financial health of companies in technology industry. Jan-Yee Kung *et. al.,* (2010) analyzes the mortgage loans of five Taiwanese commerce banks to identify the key factors that impudence prepayments and defaults, Logistic regression was used to analyze the behavior of prepayments and default. According to their study, logistic regression model was able to provide simplified results in the measurement of model variables concerning defaults and prepayments. Arun Upadhyay *et. al.,* (2012) used the Multinomial Logistic Regression (MLR) to predict the outperforming stock based on the financial ratios. Results showed high predictive accuracy rates and the model developed, can enhance an investor's stock price forecasting ability. Li et. al., (2010) used LR as a comparative method in order to build a better model for predicting stock returns effectively and efficiently. A 30 times hold-out method was used in the assessment, along with the two commonly used methods in the top 10 data mining algorithms (the support vector machine and *k* nearest neighbour) and the two baseline benchmark methods from the statistical area (MDA and LR). Sireesha (2013)examined the impact of macroeconomic factors upon the movements of the Indian stock market index Nifty, gold and silver prices through linear regression technique. Gold returns, Silver returns are selected for the analysis and are studied along with the stock returns. The performance of internal variables shows the interdependence between these variable with returns on stock, gold and silver. Stock return is significantly influenced by GDP and inflation while gold return is significantly influenced by money supply. External variables show significant impact on dependent variables.

3. RESEARCH GAP

The banking sector is significant, comparing to other sectors in stock market for local and global investors to choose various banks stocks for investment. Since Banks have major role in Indian economy, it is essential to study different factors which may influence the prices of banks stocks. Previous studies measure only the performance of stocks in general stock market. But, very few studies have been done particularly on banking sector with financial ratios to measure the performance of Banks stocks in Indian stock market. Moreover, the available literature study tells that Logistic Regression is very rarely used to build the model for forecasting the banks stock performance. Mostly, application of LR was used extensively in predicting financial distress in business failure. This study, based on financial ratios of banks, tries to develop a logistic regression model which may helpful for potential investors to look after stock performance of banks before investing.

4. RESEARCH OBJECTIVE AND METHODOLOGY

The main objective of this study is to test the model efficacy on prediction based on the financial ratios for examining the out-performing Banks in the Indian stock market.

Specific objectives are:

1. To study the effect of each financial ratio in determining the performance of the Bank's Stock.
2. To examine the applicability and appropriateness of financial ratios in the suggested Logistic regression model.

Methodology

For the purpose of the study, those Banks which have large market capitalizations have been considered and most of these Banks are part of the BSE SENSEX (i.e. BANKEX). The relevant data taken for this analysis are from websites, moneycontrol.com, Annual reports of Banks and Capitaline. The study sample consists of the ratios of 40 Banks of both nationalized and private over a period of Six-years (2010-2015), which are actively traded on the Indian stock exchange taken for classification purposes. In order to calculate the return, financial ratios and stock prices of Banks were considered.

Logistic Regression Analysis

Conceptual Framework

The logistic regression (LR) is an extended study of liner regression, can be used for studying the relationship between categorical dependent variable and multiple independent variables which may or may not be categorical. Logistic regression is based on binomial theory helpful to predict only two values.

The coefficients of the logistic regression model are coefficients like an ordinary regression equation and measures amount of variation in the dependent variable due to each of independent or explanatory variable. The coefficients are estimated not by ordinary least squares method rather, by maximum likelihood estimator (MLE) method which is a maximization of the probability of classifying the observed data into the appropriate category given the regression coefficients (Hosmer & Lemeshow, 2000).

The study uses the LR for prediction of stock performance because of the following reasons

- In LR, linearity assumption between dependent and independent variables does not exist. Rather, it is S-shaped curve.
- No need of normality assumption for independent variables is required. Also variables can be of interval scale or ratio scale or nominal scale.
- The groups (dichotomous or multinomial) must be mutually exclusive and exhaustive; a case can only be in one group and every case must be a member of one the groups.

Logistic regression is a statistical regression model for Bernoulli-distributed dependent variables and utilizes the logit as its link function.

The model takes the form

$$\text{Logit}(y) = \text{In}(\text{odds}) = \text{In}\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k.$$

where,

p = the probability of the event of interest and

$X_i = i^{\text{th}}$ predictor variable, $i = 1, 2, \ldots, k$

B_0 = constant of the equation

$\beta_i = i^{\text{th}}$ coefficient of predictor variable X_i

k = number of predictors

In LR, the odds ratios are used to understand the influence of independent variables on group membership of the outcome or event. The odds of an event are the ratio of the probability of occurrence of an event to the probability of non-occurrence of an event. The odds ratio can be interpreted as the estimated increase in the probability of success associated with a one-unit change in the value of the predictor (independent) variable.

The above model can also be written in terms of probability of an event as:

$$P_i = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k X_k}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k X_k}}$$

$P_i = E(Y = 1 \mid X)$: probability of the individual has been classified as good, given the vector X.

The estimated value of coefficient (β) has a multiplicative effect on the odds ratio.

Finally, the main purpose of logistic regression is to correctly predict the group membership of an event with the help of most parsimonious model which includes all necessary predictor variables

5. EMPIRICAL FRAMEWORK

As per the objective of the study to understand the applicability and appropriateness of Logistic regression model , first a method is required for classifying a Bank stock for investment as a "Good" or "Poor" choice for a given year. Since, there is no hard and fast rule method for classifying market investment as "Good" or "Poor", the present research used the following method which is simple and objective.

We have calculated the value of a Bank's stock and the market return for each year. For a particular year, investment option is classified as "Good" when bank's stock value rose above the market return on the other hand if it is less than market return then classified as a "Poor" investment option. Market return of the stock has been taken using the BANKEX (Index of Mumbai Stock Exchange). Market returns are calculated at the end of each financial year considering the closing prices of every year (March end).

The following formula has been used to calculate the return.

$$\text{Return of stock} = \frac{P_t - P_{t-1}}{P_{t-1}} \times 100$$

where, t = Price in the t^{th} year

P_{t-1} = Price in the $(t-1)^{th}$year

$$\text{Market return} = \frac{BEX(t) - BEX(t-1)}{BEX(t-1)}$$

The study has been done with a sample size of 40 distinct Banks' year-wise observations. In the study, dependent variable is categorical classified as "Good" or "Poor" and independent variables are 11 ratios of Banks which are shown in the table 1. These ratios are performance ratios, profitability ratios, efficiency ratios, capital adequacy ratio and equity ratio and provision ratio, which are generally, determine the value of share in the stock market.

The dependent variable (Y) based on stock market return has only two categories "Good" and "poor" and are encoded as 1 and 0 respectively.

Table 1
The eleven independent (predictors) variables are

Name of the variable	*Description of the variable*
EPS	Earnings Per share
ROA	Return on assets
ROE	Return on equity
CIR	Cost to income ratio
NIM	Net interest margin
ER	Equity ratio
CDR	Credit deposit ratio
IDR	Investment/deposit ratio
CD	Cash/deposit ratio
CAR	Capital adequacy ratio
LLP	Loan loss provision ratio

The data has been analysed through SPSS. Based on the several financial ratios considered in the study, the suggested model for the prediction of stock performance of bank is

$$\text{Logit } (y) = -7.046 - 0.003\text{EPS} - 0.462\text{ROA} + 0.254\text{ROE} + 0.060\text{CIR} - 0.387\text{NIM} + 0.221\text{ER} - 0.059\text{CDR} + 0.032\text{IDR} - 0.034\text{CD} + 0.197\text{CAR} + 0.177\text{LLP}$$

The results of the logistic regression study are analysed in four categories. Firstly, overall evaluation of the model is done. Secondly, checking the significance of each individual predictor. Thirdly, goodness-of-fit of the model is carried out using Hosmer-Lemshew statistic and finally predicted probabilities assessment has been done. Results are shown in the respective tables.

(a) The Overall Fit of a Model

Hypothesis corresponding to our study is:

H_0: Financial ratios do not affect performance of Bank's Stocks (model is not significant)

H_1: Financial ratios do affect performance of Bank's Stocks (model is significant)

The equivalent form of the above hypothesis is:

$$H_0: \beta_1 = \beta_2 = \ldots = \beta_k = 0 \text{ Vs } H_1: \text{at least one } \beta_i \neq 0$$

In order to test the above hypothesis, test statistic, -2 log likelihood is applied and which follows chi-square distribution with degrees of freedom equal to the number of predictors in the model.

The value of chi-square 224.664 shown in the Table 2,whichis significantly greater than the chi square table value and indicates the rejection of null hypothesis. Thus, study concludes that the predictors (financial ratios) are statistically significant. Table 2 also provides some approximations, *Cox and Snell's* value of 0.319 indicates that the predictors of the logistic model explains31.9% of the variation in the dependent variable. The *Nagelkerke* R square is a more reliable measure of the relationship and it ranges between 0 and 1. In our case it is 0.430, indicating a moderately good relationship of 43% between the predictors and the predicted variable.

Table 2
Model Summary

Step	*-2 Log likelihood*	*Cox & Snell R Square*	*NagelkerkeR Square*
1	224.664[a]	.319	.430

[a]Estimation terminated at iteration number 7 because parameter estimates changed by less than .001.

(b) Statistical Tests of Individual Predictors

Table 3 indicates the estimated values of coefficients and their significance in the model. (*Source:* SPSS output)

Table 3
Estimates of the coefficients of variables

		B	*S.E.*	*Wald*	*df*	*Sig.*	*Exp(B)*
Step 1[a]	EPS	–.003	.004	.550	1	.458	.997
	ROA	–.462	.898	.264	1	.607	.630
	ROE	.254	.069	13.687	1	.000	1.289
	CIR	.060	.024	6.101	1	.014	1.062
	NIM	–.387	.397	.950	1	.330	.679
	ER	.221	.224	.973	1	.324	1.247
	CDR	–.059	.027	4.616	1	.032	.943
	IDR	.032	.036	.812	1	.368	1.033
	CD	–.034	.043	.636	1	.425	.966
	CAR	.197	.164	1.438	1	.231	1.218
	LLP	.177	.091	3.770	1	.042	1.194
	Constant	–7.046	3.044	5.357	1	.021	.001

[a]Variable(s) entered on step 1: EPS, ROA, ROE, CIR, NIM, ER, CDR, IDR, CD, CAR, LLP.

The Table 3 exhibits the following:

The Wald statistic and associated probabilities which provide an index of the significance of each predictor in the model. The Wald statistic follows a chi-square distribution. It is noted from the table 3that the variables ROE, CIR, CDR, and LLP are significant ($p<0.05$) in measuring the Banks's stock performance. The LR coefficient which is statistically significant measures the impact of estimated probability and in turn the prediction of group membership.

The Exp(B) column in the table presents the extent to which raising the corresponding predictor variable by one unit influences the odds ratio. It also noted from the Table 3 that the odds ratio of independent variables ROE, CIR, ER, IDR, CIR and LLP is exceed 1 which shows that these variables will increase the occurrence of an event (good stock for investment).

(c) Goodness-of-fit Statistics

The Hosmer-lemeshow test is to examine whether the observed proportions of events are similar to the predicted probabilities of occurrence in subgroups of the model population. The test consists of grouping

the observations on the basis of expected probabilities and then testing the hypothesis that the difference between the expected and observed values is negligible or zero for all the groups. The suggested statistic follows a chi-square distribution when there is no replication in the subpopulations. The test is proposed only for binary response models. This test creates 10 ordered groups of subjects and then compares the observed numbers in each group with the predicted numbers in each group obtained by LR model. In the Table 4, the *p*-value 0.635 which is greater than 0.05, clearly indicates that there is no significant difference between observed and model-predicted values, implying that the model fits the data at an acceptable level.

Table 4
Hosmer and Lemeshow Test

Step	*Chi-square*	*Df*	*Sig.*
1	6.107	8	.635

(d) An Assessment of the Predicted Probabilities

Classification Accuracy: The classification table is a method to evaluate the predictive accuracy of the logistic regression model (Pend & So, 2002). In this, the observed values for the dependent variable and the predicted values (at a user defined cut-off value) are cross-classified. If the predicted probability of a case (response) is greater than user-specified cut off value, then it is classified into category 1, otherwise, it is classified into category 0.

The cut-off value is generally selected on the basis of

- The value which gives the best predictive fit for the available sample data (determined through trial and error.)
- The value that will separate the sample data into a specific proportion of the two groups, based on a prior known proportion split in the population.

Table 5
Classification Table[a]

		Observed	*Predicted*		
			PER		*Percentage Correct*
			POOR	*GOOD*	
Step 1	PER	POOR	107	26	80.5
		GOOD	25	72	74.2
	Overall Percentage				77.8

[a]The cut value is .500

The banks stock is classified based on the above LR model using the "*p*" values which are obtained from log odds values. For the present study, cut-off value is taken as 0.5.If the P value is higher than 0.5, then the stock is classified as good; and, if it is lower than 0.5, then the stock is classified as poor.

Table 5 presents the degree to which predicted probabilities agree with actual events (outcomes) in a classification table. In this study, 74.2 % were correctly classified for the ***Good performance* stock** group and 80.5% for the ***Poor* performance stock** group. The overall correct prediction is 77.8%. This is

a considerable improvement on the 57% correct classification with the constant model. This tells us that the model with selected predictors is a significantly better model.

6. CONCLUSION

In the present study, the LR method to predict the probability of good stock performance of bank is examined because of its potentiality to highlight the relationship between dichotomous dependent variable and one or more independent variables.

The results of the above logistic regression analysis conducted to predict the performance of banks stock with the help of different financial ratios as predictors, clearly indicates that the financial ratios do have an impact on the stock performance of the banks. Model indicates that the predictors as a set reliably distinguished between *good performance* and *poor performance* of Bank- stocks. Prediction success overall was 77.8 %(74.2% for *Good* and 80.5% for *poor*). Thus, the study reveals that the model can be used by investors, fund managers, and investment companies to enhance their ability to select out-performing Banks stock.

Scope for Further Study

The present study is based on financial ratios as the solitary factor influencing stock prices of banks, but there may be several other fiscal and organizational factors that may also influence stock prices. According to McConnell, Haslemand Gibson [1986], qualitative data can provide supplementary information to forecast stock performance more accurately. Therefore, the scope for further research lies in focusing on usage of quarterly or monthly data and diverse type of qualitative data for evaluating stock performance of Banks. The contemporary study deliberated logistic regression to construct a model, but for advance studies, a plethora of other methodologies can be considered to increase prediction ratio.

References

Avijan Dutta *et. al.,* Prediction of Stock Performance in the Indian Stock Market Using Logistic Regression, International Journal of Business and Information Volume 7, Number 1, June 2012.

Ali Ghezelbash and Farshid Keynia, A New Approach Based on Artificial Intelligent Models for Forecasting of Financial Markets, IOSR Journal of Engineering (IOSRJEN), Vol. 04, Issue 03 (March. 2014), ||V5|| PP 46-52.

Kryzanowski, L., Galler, M., & Wright, W. (2009). Using artificial neural networks to pick stocks. Financial Analyst Journal, 49, 21–27.

Mu-Yen Chen, Predicting corporate financial distress based on integration of decision tree classification and logistic regression, Expert Systems with Applications, Volume 38, Issue 9, September 2011, Pages 11261–11272.

David Booth, CEO Bonuses as Studied by Robust Logistic Regression, Journal of Data Science 01/2011; 9:293-310.

Athanasoglou P.P., Brissimis, S.N. & Delis M.D. (2008). Bank-specific, Industry-specific and Macroeconomic Determinants of Bank Profitability, International Financial Markets, Institutions & Money, Vol. 18:.121-136.

Bianco and Yohai, Robust Estimation in the Logistic Regression Model, Robust Statistics, Data Analysis, and Computer Intensive Methods, Lecture Notes in Statistics Volume 109, 1996, pp 17-34.

Sittichai Puagwatanaa and Kennedy D Gunawardana, Business Failure Prediction Model: A Case Study of Technology Industry in Thailand, Second International Conference on eLearning for Knowledge-Based Society, August 4-7, 2005, Bangkok, Thailand.

Jan-Yee Kung *et. al.,* Application of logistic regression analysis of home Mortgage loan prepayment and default risk, ICIC Express Letters, ICIC International, Volume 4, Number 2, April 2010

Arun Upadhyay, Forecasting Stock Performance in Indian Market using Multinomial Logistic Regression, Journal of Business Studies Quarterly 2012, Vol. 3, No. 3, pp. 16-39

Li, Hui, *et. al*., 2010. "Predicting business failure using classification and regression tree: An empirical comparison with popular classical statistical methods and top classification mining methods, Expert Systems with Applications 37(8), 5895-5904.

P. Bhanusireesha, effect of select macroeconomic variables on stock returns in India, International Journal of Marketing, Financial Services & Management Research, Vol. 2, No. 6, June (2013).