Stock Market Efficiency in India: An Empirical Analysis

[1]Tuhin Mukherjee, [2]Kuntal Chakraborty

Assistant Professor, Department of Business Administration, University of Kalyani, WB, 741235, India.

Research Scholar, Department of Business Administration, University of Kalyani, WB, 741235, India.

tm.ku.finance@gmail.com, kuntalwb@gmail.com

Abstract: This paper tests the Efficient Market Hypothesis in Indian stock market. An attempt has been made to discover the fifth day closing price with immediate four days closing price. Linear Regression and ARIMA (Autoregressive Integrated Moving Average) and ANN (Artificial Neural Network) are applied forecasting purpose. Out of all, linear regression gives best accuracy in case of prediction. The overall study shows, for the respective time period (17/9/2007 to 28/6/2017) Indian stock market is quite inefficient in nature. Our findings challenge the future researchers to consider other statistical tools as well as variation of number of independent variables and the overall time period to test efficiency of Indian stock market further.

Keywords: **Efficient Market Hypothesis (EMH), Artificial Neural Network (ANN), Stock Market, Linear Multiple Regression, ARIMA JEL Classification: G1, G14, G17.**

I. INTRODUCTION

The research on EMH (Efficient Market Hypothesis) is a significant topic for financial researchers through decades. Prediction of Stock price is very important for the investors who want to maximize their return and minimize their risk according to modern portfolio theory. According to EMH Financial market is very much random and efficient because of its complex dynamics behavior. There are two common methods to predict stock market, fundamental and technical approaches. Proposed method is based on technical approach that uses past form of EMH (efficient market hypothesis). Here we apply three techniques to forecast 5^{th} day price of nifty50 from immediate four-day lagged price. The all three type predicted results are compared with actual observed values for interpretation. A few statistical parameters and hypothesis test have done to compare the accuracy.

This paper is organized as follows: Section2 describes related research work, section3 introduces proposed research methods, section4 says about data set used and experimental result and representation of a graphical figure, Section5 includes the null hypothesis, Section6 ultimately draws the conclusion of this research work and Section7 highlights the references which are taken for this study.

II. RELATED WORK

Stock market forecasting is a very much challenging subject in financial field. To forecast financial time series the predictive tools have been used by taking different level of assumptions through the years. It ranges from traditional econometric methods to modern tools like artificial intelligence. Forecasting of Stock market is very much challenging task because of its non-linear, complicated, nonparametric type nature. In this context, Ghasemiyeh, Moghdani and Sana (2017) have found machine learning may be a good technique to forecast stock price. Yıldıran and Fettahoglu (2017) applied ARIMA method to predict USA exchange rate. In this case short term prediction outperforms long term prediction. Basallo, Rodriguez and Benitez (2017) applied regression technique to forecast time series for short term business problems. Erdogan (2013) has shown that in some cases a few assumptions that might not be practical could affect the forecasting accuracy. Molnar (2016) has shown the modern econometric tool of volatility measurement like GARCH type models have shown better efficiency than the traditional econometric measurement tools. Since then there are different assumption are made by different researchers therefore their prediction accuracy is also varied. Bonilla and Sepulveda (2011) has found in certain cases GARCH type of models fails to forecast stock market. Combined model shows better accuracy than individual model. Kumar and Thenmozhi (2012) has done an empirical comparison of accuracy performance by introducing different soft computing methods with varying parameters and in this case ANN shows quite better performance than others. Qiu, Song and Akagi (2016) have shown that by changing optimization parameters in case of ANN model the prediction accuracy is quietly improved in the case of Japanese stock index forecasting.

 DOI : 10.18231/2454-9150.2018.0917

III. RESEARCH METHODOLOGY

This research is focused on to discover the relation a 5th day close price with immediate 4 days close price. At first, we apply linear regression to discover 5th day closing price which is taken as dependent variable and previous 4 days closing price are independent variables. Here we have taken unadjusted predicted values of 5th close price which is resulted from linear regression. In case of ARIMA model we first check the stationary of time series by using Unit Root Test (ADF) by EVIEWS. We do the first difference of time series data to make it stationary and then apply ARIMA model using IBM SPSS. We apply this model by varying different parameters of lag(p), Difference(d), Moving average(q) values. In case of transfer function for independent variables we have taken the all parameters (Numerator, Difference, Denominator, Delay) values randomly and checked it by taking different combination for checking accuracy of ARIMA model. In all cases for transfer function we make difference of series at 1 and all delay values to 0. Here, we consider four parameters to select optimum ARIMA model which are R-Squared (Unadjusted), RMSE (Root mean square error), normalized BIC (Bayesian information criterion) and ljung-box Q statistics (for residuals) at optimum level to select optimal ARIMA model. Maximum lag length is used for ARIMA model prediction is randomly taken as 25. In case of artificial neural network(ANN) we use 5th day close price as an dependent variable and previous four day close price as co-factors and then apply multi-layer preceptor to get predicted output. We use 89% of the dataset in training and the rest on testing. Total numbers of hidden layers are two and activation functions for both hidden and output layer are hyperbolic tangent. The number of unit used in each layer is 50 at maximum. Types of training used here is batch and optimization algorithm is gradient descent. The three predicted result is compared with observed result. We use EVIEWS, IBM SPSS and MS Excel for the calculation purpose. To compare the accuracy for three methods we use the following statistical parameters:

- Descriptive test i.e. mean and standard deviation(s.d), skewness ,kurtosis
- Pearson coefficient of correlation among observed and predicted value derived from two upper said methods
- Coefficient of determination
- MAPE (mean absolute percentage error)
- RMSE (root mean square error)

Non parametric test will be used when the sample do not follow normal distribution. Here by using kolmogorov-smirnov and shapiro-wilk test it is assured that the data under test do not follow normal distribution. Thereafter to test the null hypothesis we use wilcoxon signed rank test in SPSS. The null hypothesis for this test is that "The median of the two samples that is the predicted result for three individual different techniques are significantly indifferent in nature". This test is one of the non-parametric versions of one sample or paired t test and it can be used in case of ordinal or continuous variables.

IV. DATASET

Dataset is used here collected from NSE historical data. Empirical value of Closing price of nifty50 is collected under the given sample from 17/9/2007 to 28/6/2017. The total number of data under given sample is 2415. The all data is presented (483*5) matrix format in MS Excel. Each 5th day close price is assumed is dependent for previous consecutive closing price of nifty50. Three methods for this forecast, which are linear multiple regression, ARIMA and artificial neural network (ANN) are used. Predicted values given by three methods are compared statistically with real observed values for interpretation.

V. EXPERIMENTAL RESULT

Prediction Accuracy Comparison:

Table 1: Descriptive Statistics

	Mean	Std. Deviation	Variance
DAY5	6214.67941749	1682.11736778	2829518.839
ARIMA	6215.84412937	1681.19629147	2826420.970
REGRESSION	6214.81555018	1680.25431893	2823254.576
ANN	6211.48529805	1685.84817769	2842084.078

[Source: Authors' own calculation, Software used: SPSS]

Table 2: Distributional Characteristics of Predicted Values

	Skewness		Kurtosis	
	Statistic	Std. Error	Statistic	Std. Error
DAY5	.216	.111	-.690	.222
ARIMA	.231	.111	-.702	.222
REGRESSION	.230	.111	-.700	.222
ANN	.134	.111	-.941	.222

[Source: Authors' own calculation, Software used: SPSS]

Table 3: Performance Accuracy Comparison

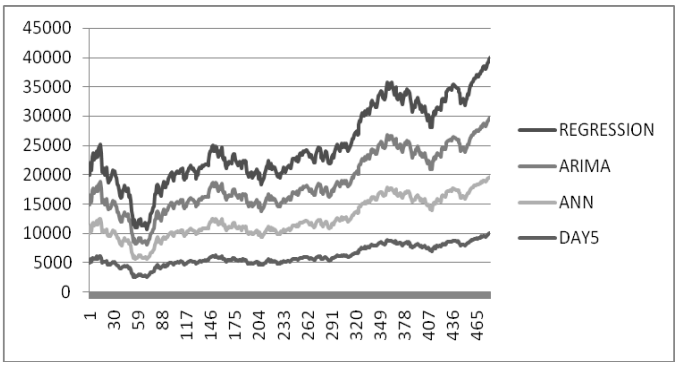
Performance Measures	ARIMA Prediction	Regression Prediction	ANN Prediction
MAPE	0.00991	0.00977	0.01650
RMSE	77.584	76.411	125.586
R	0.999	0.999	0.997
R-SQUARED(Unadjusted)	0.998	0.998	0.994

[Source: Authors' own calculation, Software used: SPSS]

In case of forecasting performance measures test parameter like MAPE (mean absolute percentage error) is more in case of ANN (artificial neural network) than in case of ARIMA and linear regression. MAPE is measured in absolute terms by neglecting sign parameters. RMSE (root mean square deviation) is better measure for accuracy in

 DOI : 10.18231/2454-9150.2018.0917

prediction. In this case also RMSE is very much lower in regression analysis than ARIMA and ANN prediction. Pearson coefficient of correlation (R) is calculated for actual closing price with the calculated forecast value of the three upper said methods. In these cases the correlation coefficient is highly positively correlated with observed value of both models but in case of regression prediction it is better than ARIMA and ANN methods. Coefficient of determination (R-squared) is taken as unadjusted value. R-squared is used to explain the proportion of sample which can be explained under testing. Although all predicted results are good but in case of Regression and ARIMA, it is 99.8% but in case of ANN it is 99.4%



[**Figure.1: Day5 observed values vs. Regression vs. ARIMA vs. ANN predicted values**]

[Source: Authors' own calculation, Software used: MS EXCEL**]**

VI. TESTING OF NULL HYPOTHESYS

Table 4:Tests of Normality for Predicted Results

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
DAY5	.141	481	.000	.951	481	.000
ARIMA	.141	481	.000	.950	481	.000
REGRESSION	.140	481	.000	.950	481	.000
ANN	.136	481	.000	.943	481	.000

[Source: Authors' own calculation, Software used :SPSS**]**

Table 5: Wilcoxon Signed Ranks Test

Test Statistics[a]			
	ARIMA - DAY5	REGRESSION - DAY5	ANN - DAY5
Z	-.302[b]	-.058[c]	-1.288[b]
Asymptotic Sig. (2-tailed)	.762	.954	.198

[Note: a. Wilcoxon Signed Ranks Test(at 5% level of significance)
b. Based on negative ranks.
c. Based on positive ranks.]
[Source: Authors' own calculation, Software used: SPSS**]**

VII. CONCLUSION

In this study we use traditional tool like linear regression and nonlinear econometric model like ARIMA to modern day soft computing tool like artificial neural network (ANN) to forecast stock market in Indian context for testing its efficiency. The findings from this study shows ANN does not show any type of superiority than linear multiple regression as well as ARIMA in terms of forecast accuracy. Linear Regression has given best predicted result by considering the all of our findings. Normality test found that no predicted value follows normal distribution. Since asymptotic significance (2 Tailed) of wilcoxon signed rank test in each case is quite higher than 0.05, we have to accept the null hypothesis and concludes that the forecasted results arrived from three different methods are quietly indifferent in nature. It is the task of future researchers to consider other statistical models and specifications as well as to change the time period and number of independent variables to test the Indian stock market efficiency further. From our overall study it would be concluded that by considering the respective data set and time period Indian stock market is quite inefficient in nature.

REFERENCES

[1] Ghasemiyeh, R., Moghdani, R., & Sana, S. S. (2017). A hybrid artificial neural network with metaheuristic algorithms for predicting stock price. Cybernetics and Systems, 48(4), 365-392.

[2] Yildiran, C. U., & Fettahoglu, A. (2017). Forecasting USDTRY rate by ARIMA method. Cogent Economics & Finance, 5(1), 1335968.

[3] Basallo-Triana, M. J., Rodriguez-Sarasty, J. A., & Benitez-Restrepo, H. D. (2017). Analogue-based demand forecasting of short life-cycle products: a regression approach and a comprehensive assessment. International Journal of Production Research, 55(8), 2336-2350.

[4] Eradogan, B. E. (2013). Prediction of bankruptcy using support vector machines: an application to bank bankruptcy. Journal of Statistical Computation and Simulation, 83(8), 1543-1555.

[5] Molnar, P. (2016). High-low range in GARCH models of stock return volatility. Applied Economics, 48(51), 4977-4991.

[6] Bonilla, C. A., & Sepulveda, J. (2011). Stock returns in emerging markets and the use of GARCH models. Applied Economics Letters, 18(14), 1321-1325.

[7] Kumar, M., & Thenmozhi, M. (2012). A hybrid ARIMA-EGARCH and Artificial Neural Network model in stock market forecasting: evidence for India and the USA. International Journal of Business and Emerging Markets, 4(2), 160-178.

[8] Qiu, M., Song, Y., & Akagi, F. (2016). Application of artificial neural network for the prediction of stock market returns: The case of the Japanese stock market. Chaos, Solitons & Fractals, 85, 1-7.

 DOI : 10.18231/2454-9150.2018.0917