Multiple Regression Prediction of Stock Intraday Prices

Calwin S. Parthibaraj
Assistant Professor
Department of Mechanical Engineering
Dr. Sivanthi Aditanar College of Engineering, Tiruchendur, Tamil Nadu, India

Abstract— In the global merchandise market, stock prices are highly volatile and strongly influenced by factors that are uncertain and indeterministic in nature. Hence, the stocks are traded with high risk by the individuals who fail to compete with the institutional investors thereby increasing the price and market volatility. To avoid the risk and to gain profit, chart tools and statistical techniques have been used. However, the techniques used by the statistician are univariate or bivariate in nature, and it neglects the consideration of multiple variables as a whole. Hence, a multiple regression analysis of variables affecting the stock prices are done in this paper using the transparent data available from National Stock Exchange (NSE). The significant factors are identified, stock price is predicted, and the confidence levels of responses are determined. The proposed methodology produced better prediction when compared with unassisted amateur traders.

Key words: Multiple Regression Analysis, Intraday Price Forecast

I. INTRODUCTION

Globalization and market liberalization has forced companies to be competitive in order to survive in the dynamic and uncertain market. In this market environment, individual traders risk and compete with the institutional traders to earn profit. Institutional traders are intellectual software traders and they are fast with heavy transactions. And individual traders are incomparable with institutional traders yet they compete and risk their money in the secondary stock market to earn profit. As the secondary stock market is affected by various tangible and intangible factors, it is necessary to determine the significant factors that forecast the stock price. Hence, the tangible predictors that are mostly scanned by traders are used to identify the significant factors to forecast. They are

1) Current gap (Last traded close price minus current period open price)
2) Open gap (Open price minus previous day close price)
3) Last traded price
4) Current traded volume
5) Market capitalization

These basic parameters are mostly used by the traders as a thumb rule without any statistical analysis to predict the response value. Moreover, individual traders face difficulties in determining significant factors and incorporating all the factors to predict stock. As each predictor has its own influence on the response variable, a multiple linear regression model is considered to determine the coefficients of the influencing factors, predict price and obtain the support/resistance line. Section 2 presents the literatures reviewed for this study, Section 3 presents the objective of the work followed by methodology, evaluation, results and discussion, and conclusion in further sections.

II. LITERATURE REVIEW

The literatures that are reviewed to explore the state of the art in stock price prediction are

Zarandi et al.(2012) presented a distributed fuzzy multi agent prediction model to handle market uncertainty in stock price predictions. The results were accurate, however the model needed extension in combining outputs into a single prediction considering social and economic factors.

Louwerse & Rothkrantz (2014) proposed an artificial neural network forecasting model to outperform naive and buy/ hold strategies. But the simulation need sustainability in determining stock price under uncertainty.

Zhai et al.(2007) combined the information from news and technical indicators to predict accurately using support vector machine. Yet the output lacked in determining the level of change in stock rise and fall.

Yeh et al. (2011) predicted stock market price by combining gradient and Lagrange multipliers to obtain weights for support vector regression method. The obtained results were better compared to trial and error method, and confirmed the use of regression in stock price prediction.

Matias & Reboredo. (2012) compared price forecasting models and found nonlinear models predicted better for sustainable time.

Lahmiri (2016) applied a new adaptive multiresolution technique with back propagation neural network to predict intraday stock prices and observed that combination of techniques would provide better results.

Babakhani et al. (2014) investigated through empirical analysis the relationship between the factors affecting earnings, and obtained positive relationship with return on assets, institutional investors and net profit.

From the literature study it is revealed that the used methodologies have been precisely modified to obtain better forecast. However the contribution of all significant independent variables was not considered as a whole in predicting stock price. Hence, the need to involve the multiple significant factors in predicting stock price is presented in next section.

III. OBJECTIVE OF THE WORK

The key objective of this paper are as follows

- Forecast the price of the trading NSE stock
- Identify the significant factors in predicting stock price
- Propose the support and resistance price level for the stock

IV. MLR METHODOLOGY

In this section, the methodology to predict the intraday stock price using multiple regression method is presented. The methodology comprises the following steps

1) Step 1: Input the stock price and the relevant factors predominantly used in its prediction
2) Step 2: Fit a Multiple Linear Regression model and obtain the coefficients for each variables
3) Step 3: Evaluate the input data to validate the Multiple Linear Regression (MLR) model assumptions and transform the data if needed to satisfy it. The MLR assumptions are
   - The mean of the response, $E(Y_i)$ at each set of values of the predictors, $(x_{1i},x_{2i},\ldots)$ is a linear function of the predictors.
   - The errors, ε_i, are Independent.
   - The errors, ε_i, at each set of values of the predictors, $(x_{1i},x_{2i},\ldots)$are Normally distributed.
   - The errors, ε_i, at each set of values of the predictors, $(x_{1i},x_{2i},\ldots)$ have Equal variances (denoted σ^2).
4) Step 4: Recalculate the regression coefficients for the full and reduced MLR model
5) Step 5: Testing the significant using general linear "F" test of each variable, and determination of partial R^2 value for the reduced model, and forecasting the stock price with its confidence levels using reduced model. These steps are presented in figure 1 as follows

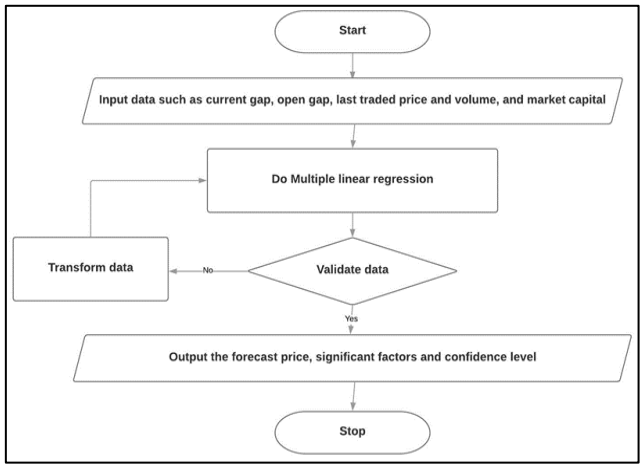

Fig. 1: Flow Chart of MLR Methodology

V. NUMERICAL EVALUATION OF MLR METHODOLOGY

The procedure determine the price forecast, confidence limits, and adjusted R^2 value to test the significance is presented in this section.

1) Step 1: The NSE data for the stock selected for intraday (NSE: KESORAMIND) is obtained from google finance and are presented in Table 1.

S.No	Stock Name	Last trade time	Current gap (₹)	Open Gap (₹)	Last Traded Price (₹)	Current Traded Volume (lacs)	Market capital (₹)
1	NSE:KESORAMIND	11:28:22 AM	3.35	1.4889	80.75	2998600	11495395096
2	NSE:KESORAMIND	2:33:58 PM	3.5	1.5556	80.9	3678294	11509630662
3	NSE:KESORAMIND	2:49:55 PM	4.35	1.9333	81.75	3778589	11637752930
4	NSE:KESORAMIND	3:03:54 PM	4.05	1.8000	81.45	3944076	11609281797
5	NSE:KESORAMIND	3:08:03 PM	4.35	1.9333	81.75	3973480	11687577954
6	NSE:KESORAMIND	3:20:01 PM	3.75	1.6667	81.15	4126523	11609281797
7	NSE:KESORAMIND	3:27:44 PM	2.75	1.2222	80.15	4211461	11438451745
8	NSE:KESORAMIND	3:30:00 PM	2.9	1.2889	80.3	4243761	11431334505

Table 1: Input Data for MLR Methodology

2) Step 2: As large number of predictors are considered, matrices are used to obtain regression coefficients and multiple linear regression model. The formula to obtain least square estimates in matrix form is given in equation (1) as

$$b = \left[b_0 \;\; b_1 \;\; b_2 \ldots . \; b_{p-1}\right]' = (X'X)^{-1}X'Y \quad (1)$$

Where the set 'b' contains the regression coefficients in transpose form, for the predictors in matrix form

$$X = \begin{bmatrix} 1 & 3.35 & 1.4889 & 2998600 & 11495395096 \\ 1 & 3.5 & 1.5556 & 3678294 & 11509630662 \\ 1 & 4.35 & 1.9333 & 3778589 & 11637752930 \\ 1 & 4.05 & 1.8 & 3944076 & 11609281797 \\ 1 & 4.35 & 1.9333 & 3973480 & 11687577954 \\ 1 & 3.75 & 1.6667 & 4126523 & 11609281797 \\ 1 & 2.75 & 1.2222 & 4211461 & 11438451745 \\ 1 & 2.9 & 1.2889 & 4243761 & 11431334505 \end{bmatrix} \quad Y = \begin{bmatrix} 80.75 \\ 80.9 \\ 81.75 \\ 81.45 \\ 81.75 \\ 81.15 \\ 80.15 \\ 80.3 \end{bmatrix}$$

The obtained multiple linear regression line of model fully considering the independent variables is given as equation (2) as

$$Y_i = -15201.32 - 23026.13 \times x_1 + 57559.06 \times x_2 + 0 \times x_3 + 0 \times x_4 + 0 \times x_5 \quad (2)$$

3) Step 3: Validate the MLR assumptions by
   - Creating a scatterplot with the residuals, e_i, on the vertical axis and the fitted values, y_i, on the horizontal axis and assess whether the residuals are nearer to zero without any outliers.
   - Creating a series of scatter plots with the residuals, e_i, on the vertical axis and each of the predictors, x_i in the model on the horizontal axes and assess each plot for near zero residuals, constant spread, and no outliers
   - Creating a histogram and normal probability plot of the residuals, e_i to check for approximate normality.
   - Checking for a strong linear or simple nonlinear trend in the resulting plot to indicate the variable plotted on the horizontal axis might be usefully added to the model

The scatter plots and the histogram is presented in figures 2.Based on the assessment, the data are not equally spreaded and grouped nearer to zero residual line. Hence, the individual data is transformed logarithmically and the step 2 is initiated for the full model with transformed data.

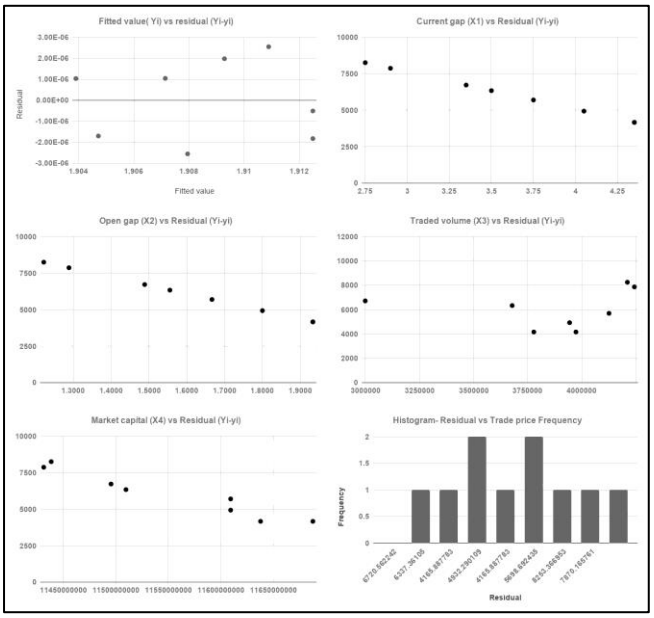

Fig. 2: Scatter Plot before Data Transformation

4) Step 4: Regression coefficients for the full and reduced MLR model

The regression coefficient are recalculated with the transformed data and the residual is plotted to check the MLR assumption. The validation of the model showed conformability of the data with the assumptions such as Linearity, normality and scatter nearer to zero residual line.

The multiple regression equation(3) thus obtained for the full model to predict the stock price after logarithmic data transformation is

$$Y_i = 1.8895 - 0.1263 \times x_1 + 0.2961 \times x_2 - 7.3 \times 10^{-5} \times x_3 + 0.00144 \times x_4 \quad (3)$$

5) Step 5: Perform general F statistic to reduce the number of insignificant variables from the full model to obtain reduced model with significant predictor variable.

The general F statistic to evaluate the significance of reduced model with the full model containing all predictor variable is given by the equation(4) and the evaluation is presented in Table 2.

$$F^{\cdot} = \left(\frac{SSE(R) - SSE(F)}{df_R - df_F} \right) \div \left(\frac{SSE(F)}{df_F} \right) \quad (4)$$

Where

- SSE(R) - the error sum of squares for the reduced models
- SSE(F) - the error sum of squares for the full models
- df_R and df_F - number of error degrees of freedom (n-p) associated with the reduced and full models
- n - number of observation
- p - number of parameters taken in MLR model

S. No	Null hypothesis	General F statistic vs F tabulated	p value for significance 0.05	Inference
1	Remove variable market capitalisation(x4)	1.14<10.13	Greater p value of 0.364	Accept Null hypothesis
2	Remove variable traded volume (x3)	10.33>10.13	Low p value of 0.048	Reject Null hypothesis
3	Remove variable open gap (x2)	72.73>10.13	Low p value of 0.0034	Reject Null hypothesis
4	Remove variable current gap(x1)	62.92>10.13	Low p value of 0.0038	Reject Null hypothesis

Table 2: General F Statistic Evaluation of Variables

As market capital (x_4) is least significant, the predictor (x_4) is not considered in the reduced model. The reduced multiple regression equation thus obtained is given in equation(5) as

$$Y_i = 1.8895 - 0.1263 \times x_1 + 0.2961 \times x_2 - 6.2 \times 10^{-5} \times x_3 \quad (5)$$

The residual for the predicted value by the reduced MLR model is plotted with the independent variable and the MLR assumptions are assessed. The validation is shown in Figure 3 and found satisfactory.

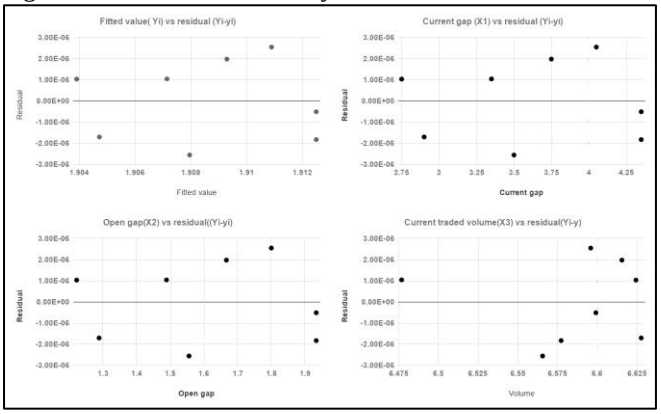

Fig. 3: Scatter Plot after Data Transformation with the Reduced Model

Further to confirm the influence of market capitalisation(x_4), the partial coefficient of determination (R^2) value is calculated to find the percent of the variation in the response by of market capitalisation(x_4). The formula used to calculate R^2 is given by equation(6)

$$R^2_{y,1,2,3|4} = \frac{SSE(reduced) - SSE(full)}{SSE(reduced)} \quad (6)$$

Using equation(6), the calculated value of R^2 proved that the three variables x_1,x_2,and x_3 explains 99.99% variation in stock price(y_h) Hence the stock price is predicted using equation(5), and the antilogarithm of predicted value is obtained to cancel the logarithmic transformations. The support and resistance level are then determined using the confidence limits for the predicted variable using equation(7) and equation(8).

The confidence limits of the response variable is given by

$$y_h \pm t_{(\alpha/2, n-p)} \times se(y_h) \quad (7)$$

$$se(y_h) = \sqrt{MSE\left(X_h^T (X^T X)^{-1} X_h\right)} \quad (8)$$

Where

- $se(y_h)$ is the standard error of response variable(y_h) given by equation(8)
- MSE is the mean square error of the predictor variables(X)
- X_h is the regression coefficient of parameters in the reduced MLR model
- $t(\alpha/2,n-p)$ is the "*t*-multiplier with "*n*-p" degrees of freedom.
- The results such as forecasted stock price and the confidence limits are obtained and presented in section 6.

VI. RESULTS & DISCUSSION

Table 3 presents the forecasted results obtained for the stock using MLR reduced model. The reduced model guarantees 99.99 percent of accuracy based on partial coefficient of determination. The reduced model also pointed positive relationship by the variable (x_2) of the stock, and could be used to predict the stock price. It is also affirmed that higher F value from general F statistic for x_2 prove its significance among other independent variables. And the support and resistance level provided strategy to take calculative risk and earn profit in dynamic and uncertain market.

S.No	Stock Name	Last trade time	Current gap (x1) (₹)	Open Gap (x2) (₹)	Current Traded Volume (lacs) (x3)	Market capital in rupees (x4)	Forecasted Price in rupees (yi)	Confidence Limits	
								Support price in rupees	Resistance Price in rupees
1	NSE:KESORAMIND	11:28:22 AM	3.35	1.4889	2998600	11495395096	80.75	68.85	94.7
2	NSE:KESORAMIND	2:33:58 PM	3.5	1.5556	3678294	11509630662	80.9	68.98	94.88
3	NSE:KESORAMIND	2:49:55 PM	4.35	1.9333	3778589	11637752930	81.75	69.71	95.87
4	NSE:KESORAMIND	3:03:54 PM	4.05	1.8	3944076	11609281797	81.45	69.45	95.52
5	NSE:KESORAMIND	3:08:03 PM	4.35	1.9333	3973480	11687577954	81.75	69.71	95.87
6	NSE:KESORAMIND	3:20:01 PM	3.75	1.6667	4126523	11609281797	81.15	69.19	95.17
7	NSE:KESORAMIND	3:27:44 PM	2.75	1.2222	4211461	11438451745	80.15	68.34	94
8	NSE:KESORAMIND	3:30:00 PM	2.9	1.2889	4243761	11431334505	80.3	68.47	94.17

Table 3: Forecasted Price and Prediction Level using MLR Methodology

VII. CONCLUSION WITH FURTHER RESEARCH

The paper presents the MLR methodology to forecast the intraday stock price. The predicted values could produce better forecast with confidence limits. It helps to prepare a trading strategy using the support and resistance level and predetermine the risk-reward ratio to earn more profits. However the MLR model needs real time implementation and evaluation for long and short positioning of stocks in traders portfolio.

REFERENCES

[1] Prempeh, K., Sekyere, A. M., & Amponsah Addy, E. K. (2018). A Multivariate Analysis of Determinants of Profitability: Evidence from Selected Manufacturing Companies Listed on the Ghana Stock Exchange.
[2] Yeh, C. Y., Huang, C. W., & Lee, S. J. (2011). A multiple-kernel support vector regression approach for stock market price forecasting. Expert Systems with Applications, 38(3), 2177-2186.
[3] Matías, J. M., & Reboredo, J. C. (2012). Forecasting performance of nonlinear models for intraday stock returns. Journal of Forecasting, 31(2), 172-188.
[4] Lahmiri, S. (2016). Intraday stock price forecasting based on variational mode decomposition. Journal of Computational Science, 12, 23-27.
[5] Tang, X., Yang, C., & Zhou, J. (2009, September). Stock price forecasting by combining news mining and time series analysis. In Web Intelligence and Intelligent Agent Technologies, 2009. WI-IAT'09. IEEE/WIC/ACM International Joint Conferences on (Vol. 1, pp. 279-282). IEEE.
[6] Zhai, Y., Hsu, A., & Halgamuge, S. K. (2007, June). Combining news and technical indicators in daily stock price trends prediction. In International symposium on neural networks (pp. 1087-1096). Springer, Berlin, Heidelberg.
[7] Zarandi, M. F., Hadavandi, E., & Turksen, I. B. (2012). A hybrid fuzzy intelligent agent- based system for stock price prediction. International Journal of Intelligent Systems, 27(11), 947-969.
[8] Louwerse, V., & Rothkrantz, L. (2014, June). Intraday stock forecasting. In Proceedings of the 15th International Conference on Computer Systems and Technologies (pp. 202-209). ACM.
[9] Ghaeli, M. (2017). Price-to-earnings ratio: A state-of-art review. Accounting, 3(2), 131-136.
[10] Selmi, N., Chaabene, S., & Hachicha, N. (2015). Forecasting returns on a stock market using Artificial

Neural Networks and GARCH family models: Evidence of stock market S & P 500. Decision Science Letters, 4(2), 203-210.

[11] Babakhani, F., Ghorbani, B., & Mohammadi, A. (2014). An investigation on important factors influencing on forecasted earnings adjustment: Evidence from Tehran Stock Exchange. Management Science Letters, 4(1), 53-56.

[12] Vazifehdust, H., Ashtiani, H. V., & Mobarhan, N. S. (2013). An application of artificial neural network to predict the added value of oil, gas and petroleum industry products. Management Science Letters, 3(8), 2349-2356.

