

---

# Interdependence of Returns on Bombay Stock Exchange Indices

**Prabhat G. Dwivedi**

Institute of Chemical Technology, Mumbai

**Ajit Kumar**

Institute of Chemical Technology, Mumbai

## ABSTRACT

*Efficient market hypothesis infers that stock market index has all information which are required for any analysis. Country's financial and economic growth reflects on stock market index and vice versa. This inspires the study on Indian stock market indices and to optimize the opportunity that Indian market gives. In this paper we studied structural interdependence on returns of some Bombay Stock Exchange indices from April 1, 2008 to March 31, 2017 using multiple linear, nonlinear tree based regression viz. decision tree, random forest and its hybrid model. We found that hybrid model gave better results as compared to multiple linear and nonlinear tree based regression.*

**Keywords:** *Multiple linear regression, Variance inflated factor, Decision Tree and Random forest.*

## 1. INTRODUCTION

Bombay stock market indices are growth measure of Indian economy, financial sectors and foreign investments. It consists of stocks from different sectors. It is also considered as portfolio bench mark for mutual funds performance and diversified portfolio. Diversified portfolio always has a problem of selecting number of stocks from different sectors. This motivates us to analyze the Bombay Stock market exchanges of six prominent indices viz. SENSEX (BSE30), BSE100, BSE200, BSE500, BSE MIDCAP (BSEMID) and BSE SMALLCAP (BSESML). There are econometric models to understand interdependence. We study the interdependence of stock indices using multi linear regression, nonlinear Tree based regression and hybrid model to understand diversification of portfolio and Indian economy. In hybrid model, we first find dependent variables by using decision tree and random forest and to these variables we fit multiple linear regression. The remaining paper is organized in five sections. We gave brief literature review in section 2. Section 3 deals with different methods used for analysis. In Section 4, we did data analysis and compared different models. Finally, we gave conclusion in section 5.

## 2. LITERATURE REVIEW

Indian economy and stock market has grown together after liberalization since 1991. If we wish to study Indian economy and its impact, we must understand the Indian stock market. (Sahu, 2016) confirmed co-integration of BSE SENSEX and macroeconomic variables like exchange rate, whole sale price index, T-bill rate and M3 since 2001 using co-integration and error correction model. (Rani, 2015) showed that stock market has negative co-integration with economic variables like exchange rate, inflation rate, index of industrial production and positive relation with money supply and yield on treasury bills. They also found that NIFTY had short run causality with exchange rate, interest rate and money supply and long run causality from inflation to Nifty and short run causality from exchange rate to Nifty. (Singh, 2016) confirmed that stock market is co-integrated by index of industrial production, inflation, money supply and exchange rate for a long run and for short run with gold price. (Ghosh., 2016) highlighted that investors attempting to diversify their investments should understand return and volatility linkages between international crude oil price, metal and Indian stock indices.

(Bharat Kolluri, 2015) found dynamic long run interrelation between India's stock and bond market which also had dynamic but not complete relation with the U.S., U.K., Japan, China and emerging Equity Markets using multivariate co-integrating tests. (Aviral Kumar Tiwari, 2015) showed that Indian stock prices are indicator for industrial growth and conclude that the economic policies should be focused with predilection on stock market environment. (Dua, 2014) examined that Foreign investment in India influences domestic equity returns, exchange rate, domestic growth and foreign output growth. (Marcelo Bianconi, 2013) showed empirically a dynamic conditional correlation between stock and bond returns from BRIC nations by using daily data from January 2003 to July 2010.

(Guidi, 2012) explored independence of Indian stock market and three developed Asian markets (i.e. Hong Kong, Japan and Singapore) and finds absence of long term relation which gives opportunity to investors in emerging markets. (Gil-Alana, 2015) suggested that the volatility of Indian National Stock Exchange (NSE) returns is persistent and asymmetric and out of sample forecasting performance is relatively poor. (Priyanka Singh, 2010) explained that index of North American, European and Asian market which close before has effect on return and volatility over other market which opens after. (Kaushik Bhattacharjee, 2014) proved bidirectional causality between the National stock exchange and NASDAQ (New York Exchange). (Bhaduri, 2011) showed "the weak correlation provides space for global funds to diversify risk in Indian markets".

### 3. METHODS

The data was divided into two parts 70% as training data and remaining 30% as testing data. We used linear and nonlinear tree based regression to understand interdependence of BSE indices on training data and verified it on testing data.

- i. **MULTI LINEAR REGRESSION:** We took return of each index and regressed with the remaining indices to find structural dependence of return of stock market indices. We observed multi co-linearity in correlation matrix which we confirmed it using Variance Inflated Factor (VIF). We dropped predictors one by one by using VIF till VIF of each variable becomes less than 5. We also checked normality of error term.
- ii. **Decision Tree:** We fitted a regression tree to understand dependence of return of index based on the remaining indices whose value was obtained by terminal nodes in the training data as the mean response of predictors return falling in that region. Thus if the test data observed to fall in that region, we made prediction with its mean value and hence concluded for its structural dependence.
- iii. **RANDOM FOREST:** Random forest regression is considered better predictor than tree regression. Hence we fitted the random forest which firstly measured variable of importance using out of bag error on each training data point and averaged over forest.
- iv. **MULTI LINEAR REGRESSION AFTER SELECTING FEATURE VARIABLES USING TREE OR RANDOM FOREST:** We observed that the decision tree or random forest had at most 3 variable of importance. Using these variables, we used multi linear regression to understand dependence response variable on explanatory variable.

### 4. DATA ANALYSIS

This study is done on secondary data of daily closing price of BSE30 (SENSEX), BSE100, BSE200, BSE500, BSEMID (MIDCAP), BESML (SMALL CAP). We downloaded this data from Bombay Stock Exchange (BSE) India website ([www.bseindia.com](http://www.bseindia.com)) for the period from April 1, 2008 to March 31, 2017. Daily closing price is highly volatile and log profit of the indices are assumed to be approximately stationary. Hence we used log return of index and referred as return. We used R software to do data analysis to understand interdependence of BSE indices.

- I. **DESCRIPTIVE DATA ANALYSIS:** Table 1 shows the descriptive analysis of the data.

**Table 1. Descriptive Data Analysis**

Statistics	Mean	Std. Dev.	Min	Q <sub>1</sub>	Median	Q <sub>3</sub>	Max
BSE30	0.030	1.532	-11.604	-0.70143	0.05479	0.71922	15.990
BSE100	0.031	1.505	-11.689	-0.69032	0.07594	0.73416	15.490
BSE200	0.032	1.474	-11.345	-0.67223	0.08971	0.73385	15.108
BSE500	0.031	1.442	-11.096	-0.66363	0.09784	0.73884	14.618
BSEMID	0.027	1.385	-8.749	-0.65010	0.12514	0.80286	11.111
BSESML	0.018	1.411	-7.972	-0.67054	0.15855	0.80153	8.660

**II. CORRELATION:** We used correlogram to find coefficient of correlation. The result is given in the Table 2.

**Table 2. Correlogram**

	BSE30	BSE100	BSE200	BSE500	BSEMID	BSESML
BSE30	1	0.99	0.99	0.98	0.83	0.75
BSE100	0.99	1	1	1	0.88	0.80
BSE200	0.99	1	1	1	0.90	0.83
BSE500	0.98	1	1	1	0.92	0.85
BSEMID	0.83	0.88	0.90	0.92	1	0.95
BSESML	0.75	0.80	0.83	0.85	0.95	1

The correlogram showed multico-linearity and strong positive correlation among returns of stock indices.

### III. STRUCTURAL DEPENDENCE OF BSE30

**A. MULTILINEAR REGRESSION:** We performed multilinear regression on BSE30 using the remaining indices and observed multico-linearity among predictor indices. We used VIF to remove multi co-linearity by dropping BSE 200, BSE 500 and BSEMID one by one as its VIF was more than 5. Finally we were left with two indices BSE100 and BSESML. We regressed BSE30 using BSE100 and BSESML and we got  $R\text{-squared} = \text{Adjusted } R\text{-squared} = 0.9908$ ,  $p\text{-value} = 2.2e^{-16}$ . We also found error to be normally distributed with  $p\text{-value} = 0.6695$  by Shapiro-Wilk normality test.

**B. TREE AND MULTILINEAR REGRESSION:** We constructed a decision tree for BSE30 and found that BSE100 was used to construct tree. Then we performed linear regression using BSE100 which gave  $R\text{-squared} = \text{Adjusted } R\text{-squared} = 0.9857$ ,  $p\text{-value} = 2.2e^{-16}$ . In this case also the error was found to be also normally distributed with  $p\text{-value} = 0.3541$  using Shapiro-Wilk normality test.

**C. RANDOM FOREST AND MULTILINEAR REGRESSION:** We constructed random forest and found two variable of importance BSE100 and BSE200. We regressed BSE30 using BSE100 and BSE200 with  $R\text{-squared} = \text{Adjusted } R\text{-squared} = 0.992$  and  $p\text{-value} = 2.2e^{-16}$ . We also observed that the error was normally distributed with  $p\text{-value} = 0.3198$  by Shapiro-Wilk normality test.

**D. COMPARISON OF MODELS USING MEAN SQUARE ERROR:** After comparing mean square error (Table 3) in various models, we found the best model was multi linear model which explained structural dependence of BSE30 with BSE100 and BSESML. The second best model was hybrid model of random forest with multi linear regression and structural dependent variable BSE100 and BSE200

**Table 3. Structural dependence of BSE30**

Model No.	Model for structural dependence of BSE30 returns	Mean square error
1	Linear regression BSE100 and BSE Small cap	0.01164233
2	Tree	0.05683731
3	Linear regression BSE100(Tree)	0.01861857
4	Random Forest	0.02795862
5	Linear regression BSE100 and BSE200 (Random forest)	0.01389838

#### IV. STRUCTURAL DEPENDENCE OF BSE100, BSE200, BSE500, BSEMID, BSESML

We performed all methods of structural dependence of BSE30 on BSE100, BSE200, BSE500, BSEMID, BSESML and got the model performance as in Table 4 to Table 8.

**Table 4. Structural dependence of BSE100**

Model No.	Model for structural dependence of BSE100 returns	Mean square error
1	Linear regression BSE30 and BSESML	0.009176429
2	Tree	0.077202809
3	Linear regression BSE200 and BSE500 (Tree)	0.002513620
4	Random Forest	0.008783273
5	Linear regression BSE200, BSE30 and BSE500 (Random forest)	0.001656326

**Table 5. Structural dependence of BSE200**

Model No.	Model for structural dependence of BSE200 returns	Mean square error
1	Linear regression BSE30 and BSESML	0.0109308131
2	Tree	0.0756403902
3	Linear regression BSE500, BSE100, BSE30(Tree)	0.0005537934
4	Random Forest	0.0021852748
5	Linear regression BSE500 and BSE100 (Random forest)	0.0005443581

**Table 6. Structural dependence of BSE500**

Model No.	Model for structural dependence of BSE500 returns	Mean square error
1	Linear regression BSE30 and BSESML	0.009450034
2	Tree	0.067445816
3	Linear regression BSE200 and BSE100(Tree)	0.001437669
4	Random Forest	0.0014344731
5	Linear regression BSE200, BSE100 and BSE30 (Random forest)	0.001318328

**Table 7. Structural dependence of BSEMID**

Model No.	Model for structural dependence of BSE30 returns	Mean square error
1	Linear regression BSE30, BSE500 and BSESML	0.06729246
2	Tree	0.22151750
3	Linear regression BSESML and BSE500(Tree)	0.009765195
4	Random Forest	0.075930981
5	Linear regression BSESML, BSE500 and BSE30(Random forest)	0.06729746

**Table 8. Structural dependence of BSESML**

Model No.	Model for structural dependence of BSESML returns	Mean square error
1	Linear regression BSE30 and BSEMID	0.1700498
2	Tree	0.3161265
3	Linear regression BSEMID(Tree)	0.1618736
4	Random Forest	0.2088422
5	Linear regression BSEMID(Random forest)	0.1618736

## V. ANALYSIS OF MODELS

We summarize the analysis different models.

- Outlier behaviour of stock market was observed as a difference among minimum, maximum, mean and median of daily return of each index. Outliers of actual data had short to long term impact on the returns of indices. Hence we did not remove the outliers.
- Random forest and decision tree gave the structural dependent variables for index with minimal number of multi-collinear variables whereas in the case of multiple regression model we had to use VIF for removal of multi co-linearity. We found that hybrid model was a better model in almost all cases with least by mean square error (MSE).
- Using multiple linear regression model, we got each of BSE100, BSE200, BSE500, BSEMID, BSESML index were dependent on BSE30 (SENSEX). Hence BSE30 influences over all Indian stock market.
- We confirmed structural dependence results with Granger Causality test.
- We also applied Bagging and Boosting but the Bagging's result was same as Tree and Boosting results were similar to Random Forest. Hence we have not given its result in detail.

## 5. CONCLUSION

In multi regression model, we found BSE30 as one of the structural variable for all other indices. This indicates that top 30 large cap company dominates the market and also decides the direction of market. This also means that a portfolio with more than 30 shares may not give better returns than BSE30. If it gives better results then return comes with high risk which can be confirmed using in mean square error. For investors, it means that a portfolio with minimum number of shares may give better return with lower risk. Non-linear model gave better order of structural variables as compared to multi linear model and corresponding hybrid model gave better static models for the index. Classifiers of decision tree can be used by practitioners and traders for their strategies of investments plans and to optimize the opportunity which stock market gives.

Limitation of this work is in the observation of outlier behaviour of indices. A separate study needs to be done to understand that. The structural dependence only gave us the long term dependence of indices. We can do further work to make a dynamic model for short term returns.

## REFERENCES

- Aviral Kumar Tiwari, M. I. (2015). Frequency domain causality analysis of stock market and economic activity in India. *International Review of Economics and Finance*(39), pp. 224-238.
- Bhaduri, S. R. (2011). Correlation dynamics in equity markets: evidence from India. *Research in International Business and Finance*, 25(1), pp. 64-74.
- Bharat Kolluri, S. W. (2015). An examination of co-movements of India's stock and government bond markets. *Journal of Asian Economics*(41), pp. 39-56.
- Dua, R. G. (2014). Foreign portfolio investment flows to India: Determinants and analysis. *World Development*, 59, pp. 16-28.
- Ghosh., S. S. (2016). Returns and volatility linkages between international crude oil price, metal and other stock indices in India: Evidence from VAR-GARCH models. *Resources Policy*, pp. 276-288.

- 
- [6] Gil-Alana, T. T. (2015). Modelling time-varying volatility in the indian stock returns: Some empirical evidence. *Review of Development Finance*, 5(2), pp. 91-97.
- [7] Guidi, R. G. (2012). Cointegration relationship and time varying co-movements among indian and asian developed stock markets. *International Review of Financial Analysis*, 21, pp. 10-22.
- [8] Kaushik Bhattacharjee, N. P. (2014). Transmission of pricing information between level ADR and their underlying domestic stocks: Empirical evidence from India. *Journal of Multinational Financial Management*, 24, pp. 43-59.
- [9] Marcelo Bianconi, J. A. (2013). BRIC and the U.S. financial crisis: An empirical investigation of stock and bond markets. *Emerging Market Review*, 14, pp. 76-109.
- [10] Priyanka Singh, B. K. (2010). Price and volatility spillovers across North American, European and Asian stock markets. *International Review of Financial Analysis*, 19(1), pp. 55-64.
- [11] Rani, D. M. (2015). Revisiting the dynamic relationship between macroeconomic fundamentals and stock prices: An evidence from indian stock market. *International Journal of Financial Management*, 5(3).
- [12] Sahu, K. K. (2016). Macroeconomic factors and the Indian stock market: Exploring long and short run relationships. *Internationa Journal of Economics and Financial Issues*, 6(3).
- [13] Singh, G. (2016, Jan). The impact of macroeconomic fundamentals on stock prices revised: A study of indian stock market. *Journal of International Economics*, 7(1), pp. 76-91.