

---

## Forecasting Indian Stock Prices and Trade Volume Using Google Trends Data

**Natchimuthu. N**

Assistant Professor

Department of Commerce, Christ University, Bengaluru.

### **Abstract:**

“Google Trends” is a public web service provided by Google Inc. It is based on search terms entered in Google search box. Google trends show how often a search term is entered in relation to the total search terms entered across the regions of the world. Google trends allow the user to compare two or more search terms. An additional feature of Google trends is the ability to link the news related to the search term in the search trend graph itself.

In this study the weekly Google trends data was used to study the relationship between the Google search trends for companies and their stock prices and stock trade volumes. The study was conducted over a time period of 11 years (2004 to 2014). Four companies from Bombay stock exchange were chosen based on their market capitalization to study the relationship. Karl Pearson’s correlation coefficient and linear regression models were used to study the relationship and the possibility of using Google trends data to predict the stock prices.

The findings show that the Google trends data has significant relationship with the stock prices of the studied companies and may possibly be used to predict the stock prices since three out of four of the linear regression models for stock prices were statistically significant. But the Google trends data for the companies does not exhibit statistically significant relationship with the stock trade volume and the linear regression models on stock trade volumes were not significant at 1% level.

**Key words:** Google trends, forecasting, linear regression models, correlation coefficient, Prediction, P-value, Significance level.

### **Forecasting Indian stock prices and trade volume using Google trends data**

#### **Introduction:**

“Google Trends” is a public web service provided by Google Inc. It is based on search terms entered in Google search box. Google trends show how often a search term is entered in relation to the total search terms entered across the regions of the world. Google trends allow the user to compare two or more search terms. An additional feature of Google trends is the ability to link the news related to the search term in the search trend graph itself.

On August 6 2008, Google launched a free service called “Insights for search”. It was meant for marketers to track various words and phrases that are entered into the Google search box. It also had the ability to categorize and organize data on geographical areas. In 2012 the Insights for search was merged with Google trends with a new interface.

#### **Statement of the problem:**

In this study Google India trends were used to forecast the stock price and trade volume for the selected BSE Sensex constituent companies.

---

### **Review of literature:**

Seth Stephens-Davidowitz(2012) used Google trends to measure variety of behaviours. He argued that the search volume data for certain words could be used to measure racism in different parts of United States.

Tobias Preis, Helen Susannah Moat, H. Eugene Stanley and Steven R. Bishop (2012) used Google trends data to show that the internet users from countries with high per capita GDP are more likely to search for information about the future than information about the past.

Tobias Preis and his colleagues Helen Susannah Moat and H. Eugene Stanley (2013) introduced a new method to identify online precursors for stock market moves, using trading strategies based on search volume data provided by Google trends.

### **Scope of the study:**

Generally the data from Google trends was used for market research. It was also used for measuring the impact of advertisement campaigns. But this study aims at looking at the possibility of using Google trends data as a predicting tool for stock prices and volume of trade.

Four companies were chosen out of BSE Sensex constituent companies based on their relevance and importance for this study.

The companies chosen were:

1. Tata Motors Ltd
2. ICICI Bank Ltd
3. Bharti Airtel Ltd
4. Tata Consultancy Services Ltd

The weekly stock prices for the selected companies were collected from Yahoo finance website.

### **Objectives:**

- To study the relationship between the Google trends data and stock prices of the study units.
- To study the relationship between the Google trends data and stock trade volume of the study units in Bombay Stock Exchange (BSE).
- To study the possibility of using Google trends data to predict the stock prices and volume of trade of the study units in BSE .

### **Period of Study:**

A period of 11 years (2004 -2014) had been chosen for this study. The Google trends data for the four selected companies had been collected from Google trends website. The weekly stock prices for the selected companies had been collected from yahoo finance website.

### **Methodology:**

This study required the analysis of Google trends data and stock prices and volume of trade for selected companies. The secondary data for the same has been collected from websites – [www.yahoofinance.com](http://www.yahoofinance.com) and [www.google.com/trends](http://www.google.com/trends). Weekly Google trends data and weekly stock prices quoted in Bombay Stock Exchange (BSE) were considered appropriate for this study. The companies chosen were selected among the BSE SENSEX constituent companies based on their market capitalization.

Google trends for a search term represent the proportion of entering a search term in relation to the total search terms entered in to the search engine. The closing stock price on Monday every week is considered as weekly stock prices.

For analyzing the possible effect of Google trends on stock prices and stock price volume, the Google trends for the first week is compared with the stock price of the following week. This method was considered appropriate to analyze the effect of the activity in Google trends on the stock prices and trade volume. To analyze the effect of Google trends on stock prices and stock trade volume, Karl Pearson’s correlation co-efficient was computed. To study the possibility of using the Google trends data to forecast the stock price and stock trade volume, Linear Regression equations were used. The calculated correlations are analyzed for significance at 1% and 5% levels. The software package SPSS was used for data analysis. The results are shown in the following pages.

**Analysis and interpretation of data:**

The data for the selected companies were analyzed for the relationship between the Google trends data and stock prices and stock trade volume. The relationship in the form of correlation coefficients between the Google Trends and the adjusted closing price and the stock trade volume were calculated. Adjusted closing share price is the price adjusted for dividends and stock splits. The results were summarized in the following pages:

**Tata Motors Ltd:**

**Correlations**

		Tata motors trends	Adjusted close	Trade volume
Tata motors trends	Pearson Correlation	1	.111*	.280**
	Sig. (2-tailed)		.010	.000
	N	535	535	535
Adjusted close	Pearson Correlation	.111*	1	.317**
	Sig. (2-tailed)	.010		.000
	N	535	535	535
Trade volume	Pearson Correlation	.280**	.317**	1
	Sig. (2-tailed)	.000	.000	
	N	535	535	535

\*. Correlation is significant at the 0.05 level (2-tailed).

\*\*. Correlation is significant at the 0.01 level (2-tailed).

The correlation coefficient observed for Tata motors shares’ adjusted closing price and Google trends for the company name is 0.111 which is significant at 5 % level and between the Google trends and trade volume for the company is 0.280 which is significant at 1%.

**ICICI Bank Ltd:**

**Correlations**

		ICICI Bank Trends	Adjusted Close	trade Volume
ICICI Bank trends	Pearson Correlation	1	.797**	.043
	Sig. (2-tailed)		.000	.300
	N	570	570	570
Adjusted Close	Pearson Correlation	.797**	1	-.305**
	Sig. (2-tailed)	.000		.000
	N	570	570	570
trade Volume	Pearson Correlation	.043	-.305**	1
	Sig. (2-tailed)	.300	.000	
	N	570	570	570

The correlation observed between Google trends and adjusted closing price is 0.797 which is significant at 1 % significance level. It shows a significant relationship between the Google search trends and the stock price quoted in the stock exchange. But the correlation between the Google trends and trade volume for ICICI Bank Ltd is not significant (.043).

**Bharti Airtel Ltd:**

**Correlations**

		Bharti Airtel Trends	Adjusted close	Trade volume
Bharti trends	Pearson Correlation	1	.468**	-.013
	Sig. (2-tailed)		.000	.752
	N	574	574	574
Adjusted close	Pearson Correlation	.468**	1	.073
	Sig. (2-tailed)	.000		.079
	N	574	575	575
Trade volume	Pearson Correlation	-.013	.073	1
	Sig. (2-tailed)	.752	.079	
	N	574	575	575

\*\* . Correlation is significant at the 0.01 level (2-tailed).

As far as Bharti Airtel Ltd is concerned, the correlation between the Google trends and adjusted closing price is significant at 1%. The relationship between the Google trends and trade volume is inverse (-0.013).

**Tata Consultancy Services Ltd :**

**Correlations**

		Tata Consultancy Services trends	Adjusted Close	Volume
Tata Consultancy Services trends	Pearson Correlation	1	-.466**	.335**
	Sig. (2-tailed)		.000	.000
	N	574	574	574
Adjusted Close	Pearson Correlation	-.466**	1	-.397**
	Sig. (2-tailed)	.000		.000
	N	574	574	574
Trade Volume	Pearson Correlation	.335**	-.397**	1
	Sig. (2-tailed)	.000	.000	
	N	574	574	574

For Tata Consultancy services Ltd, the relationship between Google trends and the stock prices is inverse (-0.466). But the correlation between the Google trends and trade volume data is significantly positive (0.335).

It is clear from the above analysis that the Google trends have a significant relationship with the stock price. Out of the four companies suited, three companies' stock prices have shown positive relationship with Google search trends. As far as the trade volumes are concerned, the relationship is positive for three out of four companies studied.

**Regression analysis:**

Regression analysis was used to check whether Google trends data can be used to explain the changes in stock prices and the share trade volume. Regression analysis was applied first keeping

Google trends as independent variable and adjusted closing stock price as a dependent variable. Then in the second half, the trade volume data was kept as dependent variable. This approach was considered appropriate to analyze the possibility of using Google trends data to forecast the stock price and trade volume. The results observed are presented in the following pages.

**Tata Motors Ltd:**

**Regression model for stock Price prediction:**

**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.111 <sup>a</sup>	.012	.010	9.6481328

a. Predictors: (Constant), trends

**ANOVA<sup>a</sup>**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	618.337	1	618.337	6.643	.010 <sup>b</sup>
	Residual	49615.086	533	93.086		
	Total	50233.423	534			

a. Dependent Variable: adjusted close

b. Predictors: (Constant), Google trends

**Coefficients<sup>a</sup>**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	17.567	.717		24.490	.000
	trends	.193	.075	.111	2.577	.010

a. Dependent Variable: adjusted close

The regression model based on Google trends explains only 1.2 % variation in Tata motors stock price(  $R^2 = 0.12$ ). In the ANOVA table it can be observed that the P value is 0.010, which shows that the regression model predicts the stock price significantly at 5% level ( $P < 0.05$ ).

**Regression model for Stock trade volume prediction:**

**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.280 <sup>a</sup>	.078	.077	788687.104

a. Predictors: (Constant), Tata motors trends

**ANOVA<sup>a</sup>**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	28215454130887.750	1	28215454130887.750	45.360	.000 <sup>b</sup>
	Residual	331540576173485.900	533	622027347417.422		
	Total	359756030304373.600	534			

a. Dependent Variable: volume

b. Predictors: (Constant), Tata motors trends

**Coefficients<sup>a</sup>**

Model		Un standardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	956329.504	58638.350		16.309	.000
	trends	41181.868	6114.589	.280	6.735	.000

a. Dependent Variable: volume

As far as Stock trade volume is concerned, the Google trends based model explains 7.8% variations in stock trade volume of Tata Motors. It can be observed from the above ANOVA table, that the model significantly predicts the trade volume for Tata Motors (  $P < 0.01$  ).

**ICICI Bank Ltd:**

**Regression model for stock Price prediction:**

**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.797 <sup>a</sup>	.635	.634	195.2314928

a. Predictors: (Constant), ICICI Bank Trends

**ANOVA<sup>a</sup>**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	37673652.515	1	37673652.515	988.412	.000 <sup>b</sup>
	Residual	21649510.725	568	38115.336		
	Total	59323163.241	569			

a. Dependent Variable: Adjusted Close

b. Predictors: (Constant), ICICI Bank trends

**Coefficients<sup>a</sup>**

Model		Un standardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	49.625	24.124		2.057	.040
	ICICI Bank	15.777	.502	.797	31.439	.000

a. Dependent Variable: Adjusted Close

For ICICI Bank, the Google trends based regression model explains 64% of the variations in the stock closing prices ( $R^2 = 0.635$ ). This regression model using the Google trends significantly predicts the ICICI Bank stock prices during the study period ( $P < 0.01$ ).

**Regression model for Stock trade volume**

**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.043 <sup>a</sup>	.002	.000	948470.844

a. Predictors: (Constant), ICICI Bank trends

**ANOVA<sup>a</sup>**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	967019875637.293	1	967019875637.293	1.075	.300 <sup>b</sup>
	Residual	510971062994204.300	568	899596941891.205		
	Total	511938082869841.600	569			

a. Dependent Variable: Volume

b. Predictors: (Constant), ICICI Bank

**Coefficients<sup>a</sup>**

Model		Un standardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	609785.418	117196.655		5.203	.000
	ICICI Bank	2527.624	2437.916	.043	1.037	.300

a. Dependent Variable: Trade Volume

The Regression model to predict the ICICI Trade volume has explained very insignificant amount of variations in stock trade volume ( $R^2 = 0.002$ ). The prediction capability of this model is not significant at even 5% level.

**Bharti Airtel Ltd:**

**Regression model for stock Price prediction:**

**Model Summary**

Model	R	R Square	Adjusted Square	Std. Error of the Estimate
1	.468 <sup>a</sup>	.219	.217	96.8878718

a. Predictors: (Constant), Bharti Airtel trends

**ANOVA<sup>a</sup>**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1503182.074	1	1503182.074	160.130	.000 <sup>b</sup>
	Residual	5369512.548	572	9387.260		
	Total	6872694.622	573			

a. Dependent Variable: adjusted close

b. Predictors: (Constant), Bharti Airtel trends

**Coefficients<sup>a</sup>**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	202.440	7.983		25.359	.000
	Bharti Airtel	2.072	.164	.468	12.654	.000

a. Dependent Variable: adjusted close

In case of Bharti Airtel, the regression model explains 22% the variations in the stock closing price ( $R^2 = 0.219$ ). The model also predicts the stock prices significantly at 1% level ( $P < 0.01$ ).

**Regression model for Stock trade volume**

**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.013 <sup>a</sup>	.000	-.002	3571052.732

a. Predictors: (Constant), Bharti Airtel trends

**ANOVA<sup>a</sup>**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1276017867785.362	1	1276017867785.362	.100	.752 <sup>b</sup>
	Residual	7294382877383368.000	572	12752417617803.092		
	Total	7295658895251153.000	573			

a. Dependent Variable: Trade volume

b. Predictors: (Constant), Bharti Airtel Trends

**Coefficients<sup>a</sup>**

Model		Un standardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	5256830.564	294229.617		17.866	.000
	Bharti Airtel trends	-1909.262	6035.780	-.013	-.316	.752

a. Dependent Variable: Trade volume

It can be observed that the regression model for predicting the Bharti Airtel stock trade volume fails to explain any variation in trade volume. The model for predicting the Bharti Airtel trade volume is not statistically significant even at 5% level ( $P > 0.05$ ).

**Tata Consultancy Services Ltd:**

**Regression model for stock Price prediction:**

**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.466 <sup>a</sup>	.217	.216	585.1147636

a. Predictors: (Constant), Tata Consultancy Services trends

**ANOVA<sup>a</sup>**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	54328494.399	1	54328494.399	158.689	.000 <sup>b</sup>
	Residual	195829511.940	572	342359.287		
	Total	250158006.340	573			

a. Dependent Variable: Adjusted Close

b. Predictors: (Constant), Tata Consultancy Services trends

**Coefficients<sup>a</sup>**

Model		Un standardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1598.817	64.797		24.674	.000
	Tata Consultancy Services trends	-33.332	2.646	-.466	-12.597	.000

a. Dependent Variable: Adjusted Close

The model for predicting stock price of Tata consultancy explains 22% of the variation in the stock price. The Regression model also predicts the stock price significantly at 1% level ( $P < 0.01$ ).

**Regression model for Stock trade volume**

**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.335 <sup>a</sup>	.112	.111	583733.701

a. Predictors: (Constant), Tata Consultancy Services trends

**ANOVA<sup>a</sup>**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	24671937159573.594	1	24671937159573.594	72.406	.000 <sup>b</sup>
	Residual	194906159470112.800	572	340745034038.659		
	Total	219578096629686.400	573			



a. Dependent Variable: Trade Volume

b. Predictors: (Constant), Tata Consultancy Services trends

**Coefficients<sup>a</sup>**

Model	Un standardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	-52888.394	64644.073		-.818	.414
1 Tata Consultancy Services trends	22461.869	2639.727	.335	8.509	.000

a. Dependent Variable: Trade Volume

The regression model for the stock trade volume explains 11% of the variation in the dependent variable i.e TCS stock trade volume ( $R^2 = 0.112$ ). It can also be observed that the model predicts statistically significantly the trade volume at 1% ( $P < 0.01$ ).

**Findings:**

The findings from the analysis of the data collected from the four companies are summarized below:

- The relationship between the Google trends data and stock prices was observed to be positive in three out of four companies studied.
- The relationship between the Google trends data and stock trade volume was also positive in three out of four companies studied.
- The linear regression models developed for predicting the stock prices were able to explain statistically significant amount of variations in the stock prices. It is ranging from 12 % in case of Tata Motors to 64% in case of ICICI Bank Ltd.
- The linear regression models developed to predict the stock trade volumes were not able to significantly explain the variations in stock trade volumes. The percentage of variations explained by the independent variable is ranging from 0% to 7.8%.
- The Linear regression models developed for stock price prediction were significant at 1% for three out of four companies studied.
- The linear regression models developed for stock trade volume prediction were significant at 5% for two out of four companies and in case of other two companies the models are not statistically significant.

**Conclusion:**

It can be concluded from the above findings, that the Google trends data for companies has significant relationship on stock prices. But in case of stock trade volumes, the Google trends data does not seem to have statistically significant relationship.

The Google trends data may be used to predict the stock prices through linear regression models. But the Google trends data fails to predict the stock trade volume for the studied companies.

**References:**

1. Tobias Preis, Daniel Reith and H. Eugene Stanley (2010). "Complex dynamics of our economic life on different scales: insights from search engine query data". *Philosophical Transactions of the Royal Society A* 368: 5707–5719. doi:10.1098/rsta.2010.0284
2. Tobias Preis, Helen Susannah Moat, H. Eugene Stanley and Steven R. Bishop (2012). "Quantifying the Advantage of Looking Forward". *Scientific Reports* 2: 350. doi:10.1038/srep00350. PMC 3320057. PMID 22482034. Retrieved April 9, 2012.

3. Paul Marks (April 5, 2012). "Online searches for future linked to economic success". New Scientist. Retrieved April 9, 2012.
4. Philip Ball (April 26, 2013). "Counting Google searches predicts market movements". Nature. Retrieved August 9, 2013.
5. Historical weekly share price data, Accessed on 21<sup>st</sup> January 2015, <https://in.finance.yahoo.com/>
6. Google trends weekly data, Accessed on 18<sup>th</sup> January 2015, <http://www.google.co.in/trends/>
7. Chris Loughlin, Erik Harnisch (2013), "The Viability of Stock Twits and Google Trends to Predict the Stock Market", [http://stocktwits.com/research/Viability-of-StockTwits-and-Google-Trends-Loughlin\\_Harnisch.pdf](http://stocktwits.com/research/Viability-of-StockTwits-and-Google-Trends-Loughlin_Harnisch.pdf).
8. Fama EF. Efficient Capital Markets: A Review of Theory and Empirical Work. The Journal of Finance. 1970 December 28-30;25(02):383-417.
9. Matthews Kayla (2014), "Google trends: How big data can Forecast the stock market", <http://www.techzone360.com/>
10. Pappas Stephanie (2013), "Google predicts stock market crashes", <http://www.livescience.com/29016-google-predicts-stock-market.html>.