WORKING PAPER 214/2022

PREDICTING POWER OF TICKER SEARCH VOLUME IN INDIAN STOCK MARKET

Ishani Chaudhuri
Parthajit Kayal

MSE
Mentoring Excellence

MADRAS SCHOOL OF ECONOMICS
Gandhi Mandapam Road
Chennai 600 025
India

February 2022

Predicting Power of Ticker Search Volume in Indian Stock Market

Ishani Chaudhuri
Madras School of Economics

and

Parthajit Kayal
(Corresponding author)
Assistant Professor, Madras School of Economics
parthajit@mse.ac.in

WORKING PAPER 214/2022

February 2022

Price : Rs. 35

MADRAS SCHOOL OF ECONOMICS
Gandhi Mandapam Road
Chennai 600 025
India

Phone: 2230 0304/2230 0307/2235 2157
Fax: 2235 4847/2235 2155
Email : info@mse.ac.in
Website: www.mse.ac.in

Predicting Power of Ticker Search Volume in Indian Stock Market

Ishani Chaudhuri and Parthajit Kayal

Abstract

This study examines the ability of online ticker searches to serve as a valid proxy for investor sentiment and forecast stock returns and trading volumes in the Indian financial market. In contrast to the common findings, we observe that ticker search volumes do not exhibit any predictive value for future excess stock returns. However, we find a weak but significant positive effect of ticker search volumes on trading volume with a two-week lag. A battery of robustness checks supports our findings. Our work warns the investors from possible misleading insights arising from search volume and stock returns related studies.

Keywords: *Online Ticker; Google Search Volume; Stock Returns; Trading Volume*

JEL Codes: *G11, G12, G15*

Acknowledgement

This paper is part of the first author's Master's thesis submitted to the Madras School of Economics. The authors would like to thank Dr. Ekta Selarka (Madras School of Economics) for her useful comments and suggestions.

Ishani Chaudhuri
Parthajit Kayal

INTRODUCTION

The famed assertion that most investors' decisions "can only be taken as a result of animal spirits of a spontaneous urge to action rather than inaction, and not as the outcome of a weighted average of benefits multiplied by quantitative probabilities” (Keynes, 1936), spurred vast amounts of research devoted to investigating the efficiency of stock markets, specifically price movements. Naturally, predicting movements of stock market variables, particularly stock returns has become one of the most heavily researched areas in finance, and presents conflicting opinions on whether it is possible to predict stock market movements. Understanding predictability (or lack thereof) of this kind can provide staggering insights into the working of financial markets.

Traditional models based on early research into financial markets assume that any new information is instantaneously integrated into asset prices. They rely on the then widely accepted efficient market hypothesis (EMH), i.e. the idea that stock prices are guided by new information as and when it becomes available, and therefore follow a random walk process (Fama, 1965). Under EMH, it is impossible to ‘beat the market’, i.e. consistently predict market movements accurately and gain abnormal returns. Later research, however, critically examines EMH and finds that stock market efficiency, if it exists, is not perfect, and prices probably cannot completely reflect all possible information (Malkiel, 2003).

There is some evidence that a predictable component exists in stock returns, but it is not easily detected (Campbell and Yogo, 2006; Cochrane 2008), and is found to be mostly limited to short horizons (Ang and Bekaert, 2007). Some studies go further to strongly reject the random walk model of stock prices (Lo and MacKinlay, 1988). Thus, the extensive literature on stock market predictability provides ample evidence both for and against the existence of efficient capital markets.
The present environment of a ‘digital age’ creates a vital opportunity for researchers to tap into previously unavailable sources of data. There is

growing recognition for the predictive value of data collected over digital platforms. Search engine traffic, i.e. frequency of an online search is one valuable repository of such data.

Google, the most popular search engine on the web, records search data for all search terms beyond a certain number of searches. This information is publicly provided and the search intensity for any keyword can be obtained through the Google Trends tool. Similarly, social media platforms such as Twitter and Facebook can also potentially provide online query data. An appealing feature of this approach is that it can provide real-time data on search behaviour in any region of the world, and has been increasingly used in recent years, from research on automobile sales (Choi and Varian, 2012), to the spread of epidemics (Carneiro and Mylonakis, 2009; Ginsberg *et. al.*, 2008; Pelat *et. al.*, 2009).

Following the intuition of Joseph *et. al.* (2011), Da *et. al.* (2011) and Bijl *et. al.* (2016), we attempt to test the validity of online ticker searches for a predefined set of firms as a proxy for the investor sentiment in the market. We use search engine data to investigate the predictive power of search engine traffic with respect to abnormal returns and trading volumes for the Indian financial market, particularly the stocks contained in the NSE Nifty 100 index as of 1st February 2020.

This is interesting as India has been lauded as the fastest-growing major economy in the past decade, only to be recently overtaken by China, and has exhibited overall growth rates of 6-9 per cent in recent years. Remarkably, the Indian economy has a rather well-developed financial market when compared to other emerging economies and has even been slightly ahead of China in this regard. Further, Indian financial markets show retail participation in large numbers as compared to developed nations (Campbell *et. al.*, 2014).

The results of the study could lead to interesting insights into how certain aspects of the stock markets in emerging economies could be very similar or very different to those in developed nations such as the United States (US).

This paper is organized as follows. In the next section, we present a detailed review of the literature. Then we describe the data and explain the methodology. Thereafter, we discuss the empirical results and robustness checks. Finally, we conclude.

LITERATURE REVIEW

There exists a suggested relationship between online search behaviour and the dynamics of market phenomena. For example, buyer behaviour theories suggest that a consumer's search for information typically precedes their purchase decision (Beatty and Smith, 1987). Over time, the impact of investor sentiment, i.e., the general prevailing attitude among investors towards the market or individual security increasingly comes into focus when studying financial markets (Barberis *et. al.*, 1998; Brown and Cliff, 2005; Baker and Wurgler, 2006). Conventional wisdom is developed, that changes in investor sentiment significantly affect contemporaneous markets. At this point, studies begin to increasingly utilize available data from news articles and headlines (Tetlock, 2007; Barber and Odean, 2008), Twitter (Bollen *et. al.*, 2011), Wikipedia (Moat *et. al.*, 2013) and Google Trends (Preis *et. al.*, 2010; Preis *et. al.*, 2013; Challet and Ayed, 2013).

In the context of financial markets, attempts over time to forecast the market conditions based on online search volumes yield mixed results. Some studies find no significant correlation between returns and search volume for stock-specific data such as company names (Preis *et. al.*, 2010), but find evidence it can instead be used to predict trading volumes. They even find that trading strategies formed based upon the search volumes can significantly dominate the market index whose constituents are under study (Preis *et. al.*, 2013; Moat *et. al.*, 2013). However, when search volumes for general financial keywords are used instead of company-specific searches the results again vary.

While some studies offer results similar to stock-specific searches (Preis *et. al.*, 2013; Moat *et. al.*, 2013), others indicate that financial terms contain no exploitable information above and beyond random, unrelated keywords regarding diseases, games, cars, etc. (Challet and Ayed, 2013). Further investigations show that portfolio diversification strategies based on search volumes also tend to outperform both the equally weighted portfolio as well as the benchmark index (Kristoufek, 2013).

Online search behaviour, particularly ticker symbol searches for different stocks can serve as a reasonable proxy for investor sentiment, (Da *et. al.*, 2011; Joseph *et. al.*, 2011). The underlying concept is that the effort required to search and analyze such data is only valuable to an individual considering investment. Further, we can plausibly rule out those looking to sell a stock as their ownership would imply former knowledge of the company's history of stock performance. Lastly, it is more likely that online ticker searches reflect the behaviour of (Joseph *et. al.*, 2011). The latter are much fewer in number have access to proprietary information that is much more accurate and would have no need to personally conduct searches of this manner containing generic information. Thus, it is logical to posit that the bulk of these searches is generated by individual retail investors who have limited influence in the market. Therefore, logical intuition doubts the real predicting power of ticker search volume.

The motivation to study the Indian stock market is threefold. Firstly, it allows us to check the reliability and predictive value of the vast, and easily obtained source of data that is google search volumes, which is freely available to all through the Google Trends tool. Secondly, to our knowledge, there is very limited literature on the subject of online ticker searches with respect to stock markets in India, particularly only the recent work of Swamy and Dharini (2019). We have a slightly different and more current sample of data. With more rigorous analysis supported by a battery of robustness tests, we contribute to a different angle to the literature on this subject for financial markets in India. Thirdly, such a study could provide powerful insights as a test of EMH for financial markets in major emerging market economies as a whole. We acknowledge that previous studies like Samanta (2004), Sarma (2004), Ahmad *et. al.* (2006) explore the EMH test on the Indian financial market. This paper complements the same literature in a slightly different context.

We investigate whether online ticker searches on company stocks can be used to predict weekly stock returns for individual firms in the Indian financial market, using the NSE Nifty 100 index and with more recent data. In keeping with empirical evidence so far, we expect the results to demonstrate that high GSV is followed by high future returns. However, our logical intuition doubts the real predicting power of ticker search volume. Therefore, this work makes an honest attempt to verify that.

DATA

The data employed in this study is obtained from CMIE ProwessIQ and Google Trends. ProwessIQ provides the daily closing prices, company betas, Nifty 100 index returns, trading volumes, dividends, and the number of shares outstanding for companies in the Nifty 100 from December 1st, 2014 through February 1st, 2020. Google search volume (GSV) data is analyzed for searches in India for the constituent stocks of Nifty 100 from December 1st, 2014 through February 1st, 2020.

The Nifty 100 index is chosen due to the relatively large size and market capitalization of the constituent companies, due to which they can be expected to have more frequent data on Google Trends. For consistency, only the companies that were in Nifty 100 in February 2020 are selected, for which complete data and GSV are available. This leaves a complete dataset on 89 companies for roughly 5 years (269 weeks). The frequency of the data is chosen to be weekly since daily data is available for very short periods of time, whereas weekly data can be collected for several years.

METHODOLOGY AND VARIABLES

EXCESS RETURNS

Excess returns and trading volume are our dependent variables for the panel regression analysis, with a focus on individual companies rather than the market as a whole, for which the effect of GSV has already been studied extensively (Challet and Ayed, 2013; Preis *et. al.*, 2013; Preis *et. al.*, 2010).

Excess returns are calculated using the capital asset pricing model. Following the work of Bijl *et. al.* (2016), first, the daily stock return for each company is calculated as total return adjusted for dividends and stock splits (1), where S is the stock closing price, D is the dividend, N is the number of shares outstanding, t is time in days, and R is the total return. From daily returns, the weekly return from the closing price of one week to the closing price of the next week is calculated (2), where n is the number of trading days in the corresponding week. From this, the weekly excess return is calculated (3) where z denotes the company, $R_{M,W}$ is the weekly market return and beta (β) is the individual beta of the company to the Nifty 100 index.

$$R_{d,t} = \frac{(S_t + D_t)N_t}{S_{t-1}N_{t-1}} - 1 \quad (1)$$

$$R_{w,t} = \prod_{i=1}^{n} R_{d,i} \quad (2)$$

$$R_{z,t} = R_{w,t} - \beta_z R_{M,W,t} \quad (3)$$

GOOGLE SEARCH VOLUME

GSV is reported by the Google Trends tool as an index of the total search volume for a particular company name over time, either globally or in specific regions. In this study, data is downloaded specific to India. Additionally, internet searches for company information within India are plausibly likely to be in reference to the Indian company rather than an alternative meaning of the word in another language, or a foreign firm with the same name. The data still suffers from considerable noise, especially for companies that sell products bearing the company name, and those that share names with other items. To circumvent this, the NSE ticker symbol for each company is used when obtaining GSV. The indices from Google Trends are used to calculate a standardized variable, SGSV (4), to make these indices more comparable. n is the number of weeks of GSV observations and σ_{GSV} is the full-sample standard deviation of the GSV time series.

$$SGSV_{z,t} = \frac{GSV_t - \frac{1}{n}\sum_{i=1}^{n} GSV_i}{\sigma_{GSV}} \quad (4)$$

TRADING VOLUME

Instead of using the raw weekly trading volume data, we include it as a detrended log volume (7), where the trend is a rolling average of the past 12 weeks of log volume. This is done to distill the deviations of trading volume from the average value at any point in time. This follows from the intuition that changes in demand that occurs slowly through time are challenging to detect using volume data due to trends in the volume itself, associated with other phenomena such as the deregulation of commissions and the growth of institutional trading (Campbell *et. al.*, 1993). We follow the intuition of previous research (e.g. Conrad *et. al.*, 1994; Cooper, 1999; Glosten *et. al.*, 1993) finds a strong connection between volume and returns and hence we include it in our model as follows:

$$Vlm_{z,t} = \log(Volume_t) - \frac{1}{12}\sum_{i=t-11}^{t} \log(Volume_i) \quad (7)$$

VOLATILITY

Weekly volatility is calculated as a square root of the sum of squared daily returns, where n is the number of trading days during the corresponding week (5). This approach assumes zero mean return in comparison to the standard deviation and makes volatility estimates more precise. (Poon and Granger, 2003). The regression model contains two measures of volatility: medium-term and long-term (Corsi, 2009). For the long-term volatility, the average of the weekly volatilities for the last five weeks is used to create a monthly variable (6).

$$\sigma_{w,z,t} = \sqrt{\sum_{i=1}^{n} r_d^2} \tag{5}$$

$$\sigma_{l,z,t} = \frac{1}{5}\sum_{i=t-4}^{t} \sigma_{w,i} \tag{6}$$

REGRESSION MODELS

Excess Returns

The explanatory variables we use are five lags each of excess return, trading volume and standardized GSV, along with short-term and long-term volatility without lags. The lags are essential so as to determine to extent to which present and past values of these variables can forecast excess returns, and how valuable the information is depending on how long ago it was available. To simplify model, we define a lag operator, L, and each explanatory variable is converted to a vector form consisting of its five most recent lags. The resulting panel regression model containing a total of 17 explanatory variables and a constant is specified as follows:

$$R_{z,t} = \alpha_0 + \left(\sum_{i=1}^{5} \beta_i L^i\right) R_{z,t} + \left(\sum_{i=1}^{5} \gamma_i L^i\right) SGSV_{z,t} + \left(\sum_{i=1}^{5} \delta_i L^i\right) Vlm_{z,t} + \rho\sigma_{w,z,t-1} + \theta\sigma_{l,z,t-1} + \epsilon_{z,t} \tag{8}$$

We conduct the Hausman specification test (Hausman, 1978) to determine whether fixed effects or random effects model is appropriate. Also known as the Durbin-Wu-Hausman test, it is a statistical hypothesis test that evaluates the consistency of an estimator when compared to an alternative estimator that is less efficient but already known to be consistent. One of the popular applications of this test for panel data is to compare the fixed effects and random effects models.

Under the null hypothesis, both models are assumed to be consistent, while the random effects model is efficient, i.e. it has minimum asymptomatic variance in the class of estimators containing the fixed effects model estimates. The alternate hypothesis is that fixed effects model is consistent while the random effects model is not. We run the same base regression for both types of models, and the difference between the model estimates, i.e. the coefficient estimate matrices (say $\beta_0 \ and \ \beta_1$), is used to calculate the Wu-Hausman statistic as follows:

$$H = (\beta_1 - \beta_0)'\big(Var(\beta_0) - Var(\beta_1)\big)(\beta_1 - \beta_0) \tag{9}$$

Under the null hypothesis, this statistic asymptotically follows chi-squared distribution with the number of degrees of freedom equal to the rank of the matrix $Var(\beta_0) - Var(\beta_1)$. If so, we accept the null and conclude that the random effects model is the appropriate one. If not, we reject the null and favour the fixed effects model.

After identifying the appropriate model, we extend it to include certain interaction effects among explanatory variables that could prove to be significant. The rationale behind this is that the predictive value of GSV for a stock at any time may somehow vary with the level of trading volume, or company's performance, managerial decisions and news surrounding it, for which excess stock returns serves as a simple proxy.

A further extension is the inclusion of the seasonal effect of the tax month, in our case, the April effect in India. Though there can be many theoretical explanations for this calendar effect (Watchel, 1942), the tax-loss selling hypothesis is the most common theory explaining this phenomenon. Individual, income-tax sensitive investors holding disproportionately small stocks, sell their holdings at the year-end, so as to claim a capital loss and reinvest at the beginning of the year. Another theory is the payment of year-end bonuses, often used to purchase stock

(Sharma and Deo, 2018). However, since the financial year in India is different and the tax year ends in March, the corresponding change would be an April effect. The resulting regression equation is as follows:

$$R_{z,t} = \alpha_0 + \left(\sum_{i=1}^{5} \beta_i L^i\right) R_{z,t} + \left(\sum_{i=1}^{5} \gamma_i L^i\right) SGSV_{z,t} + \left(\sum_{i=1}^{5} \delta_i L^i\right) Vlm_{z,t} + \eta \sigma_{w,z,t-1} + \theta \sigma_{l,z,t-1} + \lambda * April + \mu\left(R_{z,t-1} * SGSV_{z,t-1}\right) + \tau\left(Vlm_{z,t-1} * SGSV_{z,t-1}\right) + \omega\left(R_{z,t-1} * Vlm_{z,t-1}\right) + \epsilon_{z,t} \quad (10)$$

Trading Volume

We also explore the results of a similar approach for the trading volume data. Going through the same motions of conducting the Hausman specification test and using the same regressors to model trading volume. The dependent variable is simply changed to detrended log trading volume for the base model (8) and extended model (10) to assess for the impact of GSV as well as other factors on trading volume as follows:

$$Vlm_{z,t} = \alpha_0 + \left(\sum_{i=1}^{5} \beta_i L^i\right) R_{z,t} + \left(\sum_{i=1}^{5} \gamma_i L^i\right) SGSV_{z,t} + \left(\sum_{i=1}^{5} \delta_i L^i\right) Vlm_{z,t} + \rho \sigma_{w,z,t-1} + \theta \sigma_{l,z,t-1} + \epsilon_{z,t} \quad (11)$$

$$Vlm_{z,t} = \alpha_0 + \left(\sum_{i=1}^{5} \beta_i L^i\right) R_{z,t} + \left(\sum_{i=1}^{5} \gamma_i L^i\right) SGSV_{z,t} + \left(\sum_{i=1}^{5} \delta_i L^i\right) Vlm_{z,t} + \eta \sigma_{w,z,t-1} + \theta \sigma_{l,z,t-1} + \lambda * April + \mu\left(R_{z,t-1} * SGSV_{z,t-1}\right) + \tau\left(Vlm_{z,t-1} * SGSV_{z,t-1}\right) + \omega\left(R_{z,t-1} * Vlm_{z,t-1}\right) + \epsilon_{z,t} \quad (12)$$

EMPIRICAL RESULTS

HAUSMAN SPECIFICATION TEST

The results of the Hausman specification test for the base models for returns in (8) as well as trading volume in (11) reveals that the null hypotheses are rejected as the Wu-Hausman statistics are found to have statistically significant values. This means that both in the case of returns as well as trading volume, test statistic follows the chi-squared distribution asymptotically with 17 degrees of freedom. As a result, the alternative hypothesis is accepted, and the random effects model is found to be inconsistent. In other words, the fixed effects model is found to be at least as consistent as the random effects model. Therefore, we proceed with the fixed effects model for both the base models as well as extended models (10) and (12).

REGRESSION

Returns

Examining the regression results for models (8) and (10) shown below in Table 1, we find that excess returns of the week before (lag 1), two weeks before (lag 2) and four weeks before (lag 4) the current period have a consistently negative effect on present excess weekly return, significant even at the 1 percent level. This effect visibly reduces as the time between past information and current period increases. This seems to point towards a certain degree of price stickiness among stocks.

Next, the SGSV index yields insignificant results even at a 5 percent level of significance and at all five lags. Moreover, the coefficient values indicate a very weak correlation between past SGSV and excess returns. This result is not necessarily incompatible with studies that find a significant effect of GSV data when predicting excess returns (Bijl *et. al.*, 2016; Preis *et. al.*, 2013; Joseph *et. al.*, 2011), it could simply mean that the price signal received from GSV data in India, particularly for the period under consideration, is too weak to find a statistically significant result (Challet and Ayed, 2013). In fact, the negative values of the coefficients are consistent with the results of previous studies (Preis *et. al.*, 2013; Bijl *et. al.*, 2016). From our logical intuition, these results are not very surprising.

Table 1: Comparison of Models (8) and (10) for Nifty 100 Companies at Significance Levels Of 10 percent, 5 percent, and 1 percent.

Variable	Model (8)		Model (10)	
Excess return_1	-0.0405***	(-5.10)	-0.0435***	(-5.43)
Excess return_2	-0.0348***	(-4.37)	-0.0354***	(-4.45)
Excess return_3	-0.0119	(-1.50)	-0.0126	(-1.59)
Excess return_4	-0.0313***	(-3.89)	-0.0318***	(-3.95)
Excess return_5	0.00124	(0.16)	0.0012	(0.15)
sgsv_1	-0.00004	(-0.09)	-0.00002	(-0.05)
sgsv_2	-0.00002	(-0.05)	-0.00001	(-0.02)
sgsv_3	0.00054	(1.02)	0.00056	(1.05)
sgsv_4	-0.00041	(-0.77)	-0.00043	(-0.81)
sgsv_5	-0.00055	(-1.13)	-0.00055	(-1.13)
Trading volume_1	0.00015	(0.24)	0.00024	(0.39)
Trading volume_2	-0.00060	(-0.93)	-0.00059	(-0.92)
Trading volume_3	0.00092	(1.42)	0.00087	(1.35)
Trading volume_4	0.00132**	(2.05)	0.00127 **	(1.96)
Trading volume_5	-0.00024	(-0.38)	-0.00031	(-0.48)
Short term volatility	-0.178 ***	(-9.06)	-0.177***	(-9.01)
Long term volatility	0.318***	(9.74)	0.323***	(9.87)
April			0.00292**	(2.29)
Excess return_1 * sgsv_1			-0.0112	(-1.36)
Excess return_1 * Trading vol_1			0.0368***	(2.70)
sgsv_1 *Trading vol_1			-0.00039	(-0.64)
Constant	-0.00639***	(-5.93)	-0.00689***	(-6.30)

Note: * $p<0.10$, ** $p<0.05$, *** $p<0.01$

The lack of predictability can possibly be explained by breaking down the composition and behaviour of stock market investors in India. The relative participation of retail investors in India is extremely low in terms of trading volumes when compared to that of the few institutional investors. The latter have large resources to invest and trade with, and proprietary sources of accurate information regarding stock markets at their disposal to do so profitably. This, in congruence with the risk-averse tendencies of Indian retail investors to gravitate towards traditional forms of investment with their excess funds, like bank deposits, post office savings, etc. leave the Indian stock market to be dominated driven by foreign and domestic institutional investors. Other reasons for these results could include the distorting effects of noise traders and low risk appetite of retail investors in Indian stock market.

We find that for both models, the effect of detrended log trading volume is significant only at the fourth lag, and at a 5 percent level of significance. We also notice that overall, there is an alternate positive and negative effect of trading volume in successive weeks. The effect of volatility appears to be much stronger than all other regressors, with a negative effect for short term (past week) volatility and a positive effect for long term (past 5 weeks) volatility.

We see that the same patterns hold for both models (8 and 10), and so, the addition of interaction variables and an April dummy does not greatly affect the behaviour of the regressors, with negligible changes in coefficients. The April dummy has a weak but significant effect on excess returns. This is also consistent with the work of Bijl *et. al.* (2016), who employed the January effect due to working on US data and found the seasonal dummy variable to be significant. Among the interaction variables, we find that the interaction of lagged excess returns and lagged trading volume is significant, which suggests that higher excess returns in the previous week increases the predictive value of detrended log trading volume in the previous week by 36.8 basis points.

We do not report R^2 and adjusted-R^2 values for the above or any further regressions due to the values being extremely low (<1 percent). This is not surprising as it is quite common for studies of this nature to yield very low R^2 values. Given the fact that a host of macroeconomic, and company specific factors such as the union budget, GDP, inflation, money supply, working capital, foreign investment, the performance of other firms, etc. as well as changes in regulations or operation of governing bodies drive stock returns far more strongly, this is within expectation.

Past studies strongly indicate that stock returns and trading volume can be contemporaneously related, or at the very least can be highly correlated (Smirlock and Starks, 1988; Lee and Swaminathan, 2000). This can potentially lead to a problem of multicollinearity in our model. To address this concern, we check our selected regressors for correlation. We find a correlation coefficient of only 0.0026 between excess returns and detrended trading volume. Further, we check the pairwise correlation coefficient for all regressors and find all to be well below 0.5, exceeding it only slightly in the case of different lags of the same variable, which is expected (see Appendix II. It is made available upon request). We run the same regressions as in Table 1 in excluding all lags of detrended log trading volume as shown above in Table 2 and find very similar results.

Thus, according to our findings, there is no visible contemporaneous relationship between trading volume and returns (since we use the same lags for both variables), and this is corroborated in the work of Ratner and Leal (2001), who examine the Latin American and Asian financial markets and find a positive contemporaneous relation between return and volume in these countries, with the exception of India. Therefore, the investors should be aware that predicting power of search volume is very limited in the equity markets (especially in major financial markets) unlike cryptocurrencies.

Table 2: Comparison of models (8) and (10) for Nifty 100 Companies Excluding Detrended Log Trading Volumes

Variable	Model (8)		Model (10)	
Excess return_1	-0.0361 ***	(-4.63)	-0.0377 ***	(-4.81)
Excess return_2	-0.0308 ***	(-3.94)	-0.0312 ***	(-4.00)
Excess return_3	-0.00960	(-1.23)	-0.00988	(-1.26)
Excess return_4	-0.0288 ***	(-3.67)	-0.0289 ***	(-3.69)
Excess return_5	0.00430	(0.56)	0.00434	(0.56)
sgsv_1	0.00024	(0.49)	0.00022	(0.46)
sgsv_2	-0.00036	(-0.68)	-0.00036	(-0.68)
sgsv_3	0.00046	(0.87)	0.00049	(0.93)
sgsv_4	-0.00010	(-0.19)	-0.00010	(-0.19)
sgsv_5	-0.00065	(-1.35)	-0.00065	(-1.36)
Short term volatility	-0.187 ***	(-9.53)	-0.186 ***	(-9.50)
Long term volatility	0.337 ***	(10.39)	0.339 ***	(10.44)
April			0.00292**	(2.29)
Excess return_1 * sgsv_1			-0.0126	(-1.56)
Constant	-0.00666 ***	(-6.19)	-0.00703***	(-6.45)

Note: * p<0.10, **p<0.05, ***p<0.01

Trading Volume

The regression results for models (11) and (12) are reported below in Table 3. Similar to our results for returns, we find some lags of excess returns and trading volume are significant (lags 4 and 5). This is expected as the regressors remain the same as models (8) and (10) and we have already shown above that trading volume and excess returns share a statistically significant relationship.

In the earlier regressions we found that the fourth lag of trading volume was significant in predicting excess returns, so it is intuitive that we also find the fourth lag of excess returns to be significant in predicting trading volume as the data used in both regressions is the same. The fifth lag is also found to be significant at the 10 percent level of significance and we observe the same diminishing and alternately positive and negative coefficients as before.

Further, the lags of the trading volume itself are found to be significant. This too is within expectation due to a small degree of serial correlation remaining after detrending the variable, which is easily eliminated by differencing and does not change the results (we have the results in Appendix III Table A1. Appendix is available upon request).

Table 3: Comparison of Models (11) and (12) for Nifty 100 Companies at Significance Levels of 10 percent, 5 percent, and 1 percent.

Variable	Model (11)		Model (12)	
Excess return_1	0.0527	(0.53)	0.0783	(0.79)
Excess return_2	-0.0482	(-0.49)	-0.0312	(-0.31)
Excess return_3	0.114	(1.15)	0.127	(1.29)
Excess return_4	0.208**	(2.07)	0.212**	(2.12)
Excess return_5	-0.192*	(-1.93)	-0.189*	(-1.91)
sgsv_1	-0.00135	(-0.22)	-0.00242	(-0.39)
sgsv_2	0.0133**	(1.99)	0.0122*	(1.82)
sgsv_3	-0.00585	(-0.87)	-0.00603	(-0.90)
sgsv_4	-0.00184	(-0.27)	-0.00192	(-0.29)
sgsv_5	-0.00108	(-0.18)	-0.00069	(-0.11)
Trading volume_1	0.153***	(19.11)	0.149***	(18.70)
Trading volume_2	0.0582***	(7.20)	0.0578***	(7.16)
Trading volume_3	0.0240***	(2.96)	0.0259***	(3.19)
Trading volume_4	-0.0153	(-1.89)	-0.0121	(-1.50)
Trading volume_5	-0.0273***	(-3.43)	-0.0251***	(-3.16)
Short term volatility	1.174***	(4.79)	1.180***	(4.82)
Long term volatility	0.492	(1.21)	0.311	(0.76)
April			-0.121***	(-7.63)
Excess return_1 * sgsv_1			-0.0314	(-0.31)
Excess return_1 * Trading vol_1			-0.176	(-1.04)
sgsv_1 *Trading vol_1			0.00689	(0.91)
Constant	-0.0454***	(-3.37)	-0.0279**	(-2.05)

Note: * p<0.10, **p<0.05, ***p<0.01

What is interesting is that in case of both our base (11) as well as extended (12) models, we find the second lag of SGSV to be significant in forecasting detrended trading volumes, while earlier we found none of the SGSV variables to significantly predict excess returns. This is consistent with the findings of Sanchez (2020) who observes no significant influence of GSV on asset returns but report a significant effect on trading volumes for the following week. This shows that GSV data contain much stronger signals towards trading volumes rather than predicting returns.

Robustness Check

To check for the consistency of the above results, particularly the insignificance SGSV variables, we run the same panel models with fixed effects for a popular subset of the Nifty 100 companies, the Nifty 50 companies. The appeal of this somewhat oversimplified approach (as opposed to random sampling) lies in the fact that the Nifty 50 companies have the maximum market capitalisation among Nifty 100 companies, and hence it is intuitive to assume that they would account for a fairly large percentage of online ticker searches for Nifty 100 companies. We do this for the model for excess returns and obtain the following results:

Table 4: Comparison of Models (8) and (10) for Nifty 50 companies

Variable	Model (8)		Model (10)	
Excess return_1	-0.0402***	(-3.32)	-0.0421 ***	(-3.48)
Excess return_2	-0.0560***	(-4.65)	-0.0572 ***	(-4.75)
Excess return_3	-0.0112	(-0.93)	-0.0118	(-0.97)
Excess return_4	-0.0442***	(-3.58)	-0.0452 ***	(-3.66)
Excess return_5	-0.0124	(-1.00)	-0.0136	(-1.10)
sgsv_1	0.00089	(1.15)	0.00084	(1.09)
sgsv_2	-0.00014	(-0.17)	-0.00013	(-0.16)
sgsv_3	0.00094	(1.13)	0.00095	(1.13)
sgsv_4	-0.00096	(-1.15)	-0.00096	(-1.15)
sgsv_5	-0.00094	(-1.22)	-0.00092	(-1.20)
Trading volume_1	0.00030	(0.31)	0.00043	(0.43)
Trading volume_2	-0.00075	(-0.74)	-0.00075	(-0.74)
Trading volume_3	0.00217***	(2.13)	0.00214 ***	(2.10)
Trading volume_4	0.00137	(1.35)	0.00129	(1.27)
Trading volume_5	0.00066	(0.66)	0.00053	(0.54)
Short term volatility	-0.138***	(-4.79)	-0.137 ***	(-4.76)
Long term volatility	0.385***	(7.89)	0.390 ***	(7.95)
April			0.00348	(1.73)
Excess return_1 * sgsv_1			-0.0225	(-1.72)
Excess return_1 * Trading vol_1			0.0492 **	(2.38)
sgsv_1 *Trading vol_1			0.00039	(0.37)
Constant	-0.0113 ***	(-6.51)	-0.0118 ***	(-6.72)

Note: * p<0.10, **p<0.05, ***p<0.01

We note that the coefficient values seem to be consistently larger than the Nifty 100 coefficients (Table 1) for both models, sometimes as much as by 200 basis points as in the case of lagged excess returns. Further, the third lag for detrended log trading volume seems to be significant here instead of the fourth. Other than these minor differences that are well within expectation, we do not find any major changes in the regression results for a subset of the Nifty 50 companies and the results are consistent. Thus, we can conclude that the model is a consistent one.

Another robustness check we perform for the excess return models is to include an element of a study by Swamy and Dharani (2019) who follow the methodology and build on the work of Bijl *et. al.* (2016). We replace the SGSV regressor with the first difference or delta versions of the SGSV index as below in Table 5. The motivation behind this approach was to perhaps isolate the predictive element in SGSV data while eliminating any serial correlation, as is common with time series data.

Table 5: Comparison of Models (8) and (10) for Nifty 100 Companies Using Delta SGSV instead of SGSV

Variable	**Model (8)**		**Model (10)**	
Excess return_1	-0.0397 ***	(-4.98)	-0.0419 ***	(-5.24)
Excess return_2	-0.0326 ***	(-4.10)	-0.0332 ***	(-4.18)
Excess return_3	-0.00909	(-1.14)	-0.00974	(-1.23)
Excess return_4	-0.0315 ***	(-3.92)	-0.0318 ***	(-3.97)
Excess return_5	-0.00176	(-0.22)	-0.00180	(-0.22)
Delta sgsv_1	0.00011	(0.24)	0.00014	(0.31)
Delta sgsv_2	0.00021	(0.41)	0.00024	(0.47)
Delta sgsv_3	0.00086	(1.59)	0.00090	(1.65)
Delta sgsv_4	0.00051	(0.99)	0.00054	(1.03)
Delta sgsv_5	0.00006	(0.14)	0.00007	(0.15)
Trading volume_1	0.00012	(0.19)	0.00021	(0.34)
Trading volume_2	-0.00072	(-1.12)	-0.00072	(-1.11)
Trading volume_3	0.00085	(1.31)	0.00079	(1.23)
Trading volume_4	0.00135**	(2.08)	0.00129 **	(1.99)
Trading volume_5	-0.00032	(-0.50)	-0.00039	(-0.62)
Short term volatility	-0.181***	(-9.22)	-0.181 ***	(-9.22)
Long term volatility	0.321 ***	(9.84)	0.327***	(9.99)
April			0.00308**	(2.42)
Excess return_1 * Delta sgsv_1			-0.00347	(-0.36)
Excess return_1 * Trading vol_1			0.0358 **	(2.63)
Delta sgsv_1 *Trading volume_1			0.00052	(0.66)
Constant	-0.00637 ***	(-5.92)	-0.00688***	(-6.30)

Note: * p<0.10, **p<0.05, ***p<0.01

The results are again consistent with the original models, with minute differences in coefficient values and no new significant variables. Thus, we can conclude that the returns models (8) and (10) are robust to changes in sampling and that the insignificance of SGSV as a regressor at all five lags is not due to serial correlation within SGSV.

We conduct the same robustness checks for the results of models (11) and (12) for trading volume as described in Table 3, and observe similar effects, with minute changes in coefficient values and significance of different lags of regressors but with our overall results unchanged. We report these tables in the appendix rather than in this section due to the paucity of space and similar nature of the results (See Appendix III Tables A2 and A3, available upon request).

CONCLUSION

In this study, we tried to model the predictive value of google search volumes for stock returns and the trading volume for NSE Nifty 100 companies over a period of roughly five years. In contrast to the common findings, we observe that ticker search volumes do not exhibit any predictive value for future excess stock returns. Our results are similar to the findings of Challet and Ayed (2013). Further, we find a weak but significant positive effect of ticker search volumes on trading volume with a two-week lag. A battery of robustness checks supports our findings. These results confirm that the Indian financial market is efficient. Our work warns the investors from possible misleading insights arising from search volume and stock returns related studies. They should be aware that predicting power of search volume is very limited in the equity markets (especially in major financial market like India) unlike cryptocurrencies. The lack of predictability can possibly be explained by breaking down the composition and behaviour of stock market investors in India, analysing their risk averse tendencies and accounting for noise traders.

The considerably low relative participation of retail investors in India compared to that of the few institutional investors can be the primary cause. They lack the resources as well as the proprietary sources of accurate information to invest in stock markets. This, in congruence with the risk-averse tendencies of Indian retail investors to gravitate towards traditional forms of investment leaves the stock market to be dominated and driven by institutional investors.

Ticker searches are largely generated by amateur retail investors and contain a great deal of participation of noise traders. So, not only are the search volumes largely incomplete due to not being used by institutional investors, they are also full of inaccurate information about the stock market. On the other hand, experienced and trained traders may not be conducting online ticker searches. They may exhibit their sentiment towards the market in a different way such as searching company names, other financial phrases/keywords such as "currently rising stocks", following other sources of information such as the rumours and news surrounding a listed company, or suggestions of other experienced traders rather than conducting direct research in the manner of online ticker search.

Further explanations may be rooted in the risk-averse nature of Indian investors. It has been observed that in India, stocks performing well and giving higher returns tend to continue to do so, while those that underperform continue to give lower returns, and this could be due to the unwillingness of investors to invest in stocks they consider 'risky' and 'low return' in an emerging market.

Finally, Indian retail investors in the stock market tend to prefer the long term investments to short term trading in the daily or weekly horizon, and hence our analysis could suffer from the limitation of sample size. This has occurred due to the user-unfriendly restrictions in the Google Trends tool which does not allow for downloading weekly horizon data for a period longer than 5 years. There is considerable scope for verification and validation of our results using more innovative and resonant approaches, particularly while modelling trading volume.

The results of this study could have important implications for the various stakeholders in Indian financial markets. For retail investors, it serves as a warning for the possible misleading insights from ticker search analysis of stocks and resulting in unexpected losses in their portfolios. For institutional investors, it simply reinforces the value of proprietary non-public information available to them. For policymakers, it could aid in designing better monetary policies keeping in mind the implied efficiency of the Indian financial market. Lastly, in academia, it aims to encourage further study of this market on the lines of informational efficiency and stock market behaviour in this unique economy.

REFERENCES

Ahmad, K. M., Ashraf, S., and Ahmed, S. (2006), "Testing Weak Form Efficiency for Indian Stock Markets", *Economic and Political Weekly*, 49-56.

Ang, A., and Bekaert, G. (2007), "Stock Return Predictability: Is it There?", *Review of Financial Studies*, *20*(3), 651-707.

Baker, M., and Wurgler, J. (2006), "Investor Sentiment and the Cross-section of Stock Returns", *Journal of Finance*, *61*(4), 1645-1680.

Barber, B. M. and Odean, T. (2008), "All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors", *Review of Financial Studies* 21, 785–818.

Barberis, N., Shleifer, A., and Vishny, R. (1998), "A Model of Investor Sentiment", *Journal of Financial Economics*, *49*(3), 307-343.

Beatty, S. E. and Smith, S. M. (1987), "External Search Effort: An Investigation Across Several Product Categories", *Journal of Consumer Research* 14, 83–95.

Bijl, L., Kringhaug, G., Molnár, P., and Sandvik, E. (2016), "Google Searches and Stock Returns", *International Review of Financial Analysis*, *45*, 150-156.

Bollen, J., Mao, H., and Zeng, X. (2011), "Twitter Mood Predicts the Stock Market", *Journal of Computational Science*, *2*(1), 1-8.

Brown, G.W. and Cliff, M. T. (2005), "Investor Sentiment and Asset Valuation", *Journal of Business* 78, 405–440.

Campbell, J. Y., and Yogo, M. (2006), "Efficient Tests of Stock Return Predictability", *Journal of Financial Economics*, *81*(1), 27-60.

Campbell, J. Y., Grossman, S. J., and Wang, J. (1993), "Trading Volume and Serial Correlation in Stock Returns", *Quarterly Journal of Economics*, *108*(4), 905-939.

Campbell, J. Y., Ramadorai, T., and Ranish, B. (2014), "Getting Better or Feeling Better? How Equity Investors Respond to Investment Experience (No. w20000)", *National Bureau of Economic Research*.

Carneiro, H. A., and Mylonakis, E. (2009), "Google Trends: A Web-based Tool for Real-time Surveillance of Disease Outbreaks", *Clinical Infectious Diseases*, *49*(10), 1557-1564.

Challet, D., and Ayed, A. B. H. (2013), "Predicting Financial Markets with Google Trends and Not so Random Keywords", *arXiv preprint arXiv:1307.4643*.

Choi, H., and Varian, H. (2012), "Predicting the Present with Google Trends", *Economic Record*, *88*, 2-9.

Cochrane, J. H. (2008), "The Dog that did not Bark: A Defense of Return Predictability", *Review of Financial Studies*, *21*(4), 1533-1575.

Conrad, J. S., Hameed, A., and Niden, C. (1994), "Volume and Autocovariances in Short-horizon Individual Security Returns", *Journal of Finance*, *49*(4), 1305-1329.

Cooper, M. (1999), "Filter Rules Based on Price and Volume in Individual Security Overreaction", *Review of Financial Studies*, *12*(4), 901-935.

Corsi, F. (2009), "A simple Approximate Long-memory Model of Realized Volatility", *Journal of Financial Econometrics*, *7*(2), 174-196.

Da, Z., Engelberg, J., and Gao, P. (2011), "In Search of Attention", *Journal of Finance*, *66*(5), 1461-1499.

Fama, E. F. (1965). "The Behavior of Stock-market Prices", *Journal of Business*, *38*(1), 34-105.

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., and Brilliant, L. (2009), "Detecting Influenza Epidemics using Search Engine Query Data", *Nature*, *457*(7232), 1012-1014.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993), "On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks", *Journal of Finance*, *48*(5), 1779-1801.

Hausman, J. A. (1978), "Specification Tests in Econometrics", *Econometrica: Journal of the Econometric Society*, 1251-1271.

Joseph, K., Wintoki, M. B., and Zhang, Z. (2011), "Forecasting Abnormal Stock Returns and Trading Volume using Investor Sentiment: Evidence from Online Search", *International Journal of Forecasting*, *27*(4), 1116-1127.

Keynes, J. M. (1936), "The General Theory of Interest, Employment, and money"

Kristoufek, L. (2013), "Can Google Trends Search Queries Contribute to Risk Diversification?", *Scientific Reports*, *3*, 2713.

Lee, C. M. C. and Swaminathan, B. (2000), "Price Momentum and Trading Volume" *Journal of Finance*, 55(5), 2017-2069.

Lo, A. W., and MacKinlay, A. C. (1988), "Stock market Prices do not Follow Random Walks: Evidence from a Simple Specification Test", *Review of Financial Studies*, *1*(1), 41-66.

Malkiel, B. G. (2003), "The Efficient Market Hypothesis and its Critics", *Journal of Economic Perspectives*, *17*(1), 59-82.

Moat, H. S., Curme, C., Avakian, A., Kenett, D. Y., Stanley, H. E., and Preis, T. (2013), "Quantifying Wikipedia Usage Patterns Before Stock Market Moves", *Scientific Reports*, *3*, 1801.

Pelat, C., Turbelin, C., Bar-Hen, A., Flahault, A., and Valleron, A. J. (2009), "More Diseases Tracked by using Google Trends", *Emerging Infectious Diseases*, *15*(8), 1327.

Poon, S. H., and Granger, C. W. (2003), "Forecasting Volatility in Financial Markets: A review", *Journal of Economic Literature*, *41*(2), 478-539.

Preis, T., Moat, H. S., and Stanley, H. E. (2013), "Quantifying Trading Behavior in Financial Markets Using Google Trends", *Scientific Reports*, *3*, 1684.

Preis, T., Reith, D., and Stanley, H. E. (2010), "Complex Dynamics of Our Economic Life on Different Scales: Insights from Search Engine Query Data", *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 368(1933), 5707-5719.

Ratner. M. and Leal R. P. C., (2001), "Stock Returns and Trading Volume: Evidence from the Emerging Markets of Latin America and Asia", *Journal of Emerging Markets* 6(1), 5-22.

Samanta, G. P. (2004), "Evolving Weak-form Informational Efficiency of Indian Stock Market" *Journal of Quantitative Economics*, 2(1), 66-75.

Sanchez, M. G. (2020), "Investor Attention Using the Google Search Volume Index – Impact on Stock Returns" *Review of Behavioral Finance*, Vol. ahead-of-print No. ahead-of-print. https://doi.org/10.1108/RBF-01-2020-0011.

Sarma, S. N. (2004), "Stock Market Seasonality in an Emerging Market", *Vikalpa*, 29(3), 35-42.

Sharma, A. and Deo, V. (2014), "Seasonal Anomalies in Indian Stock Markets" *International Research Journal of Finance and Economics*, 118, 74-101

Smirlock, M. and Starks, L. (1988), "An Empirical Analysis of the Stock Price-Volume Relationship", *Journal of Banking and Finance*, 12(1), 31-41.

Swamy, V. and Dharini, M. (2020), "Investor Attention using the Google Search Volume Index – impact on stock returns", *Review of Behavioral Finance*, 11(1), 55-69

Tetlock, P. C. (2007), "Giving Content to Investor Sentiment: The role of Media in the Stock Market", *Journal of Finance*, *62*(3), 1139-1168.

Wachtel, S. B. (1942), "Certain Observations on Seasonal Movements in Stock Prices", *Journal of Business of the University of Chicago*, *15*(2), 184-193.

MSE Monographs

- * Monograph 34/2015
  Farm Production Diversity, Household Dietary Diversity and Women's BMI: A Study of Rural Indian Farm Households
  Brinda Viswanathan
- * Monograph 35/2016
  Valuation of Coastal and Marine Ecosystem Services in India: Macro Assessment
  K. S. Kavi Kumar, Lavanya Ravikanth Anneboina, Ramachandra Bhatta, P. Naren, Megha Nath, Abhijit Sharan, Pranab Mukhopadhyay, Santadas Ghosh, Vanessa da Costa and Sulochana Pednekar
- * Monograph 36/2017
  Underlying Drivers of India's Potential Growth
  C.Rangarajan and D.K. Srivastava
- * Monograph 37/2018
  India: The Need for Good Macro Policies *(4th Dr. Raja J. Chelliah Memorial Lecture)*
  Ashok K. Lahiri
- * Monograph 38/2018
  Finances of Tamil Nadu Government
  K R Shanmugam
- * Monograph 39/2018
  Growth Dynamics of Tamil Nadu Economy
  K R Shanmugam
- * Monograph 40/2018
  Goods and Services Tax: Revenue Implications and RNR for Tamil Nadu
  D.K. Srivastava, K.R. Shanmugam
- * Monograph 41/2018
  Medium Term Macro Econometric Model of the Indian Economy
  D.K. Srivastava, K.R. Shanmugam
- * Monograph 42/2018
  A Macro-Econometric Model of the Indian Economy Based on Quarterly Data
  D.K. Srivastava
- * Monograph 43/2019
  The Evolving GST
  Indira Rajaraman

MSE Working Papers

Recent Issues

* Working Paper 206/2021
  Integration of Econometric Models and Machine Learning- Study on US Inflation and Unemployment
  Sri Rajitha Tattikota and Naveen Srinivasan

* Working Paper 207/2021
  Gender Parity in Higher Education Enrolment: Role of Family Networks
  Kavya Ravindranath and Brinda Viswanathan

* Working Paper 208/2021
  Gender Differences in Double Burden of Malnutrition in India: Quantile Regression Estimates
  Archana Agnihotri and Brinda Viswanathan

* Working Paper 209/2021
  Essay on Non-linear Pricing in E-commerce
  Dipankar Das and Vivek Sharadadevi Jadhav

* Working Paper 210/2021
  Role of ICT Dissemination and Digital Finance in Poverty Eradication and Income Inequality Reduction: A Sub-national Level Study from India
  Simontinti Das and Amrita Chatterjee

* Working Paper 211/2021
  Are India's farm debt waivers a political tool that impacts government finances?
  Sowmya Dhanaraj, Vidya Mahambare and Pragati

* Working Paper 212/2021
  Rural Urban Differentials in Health Insurance Demand
  Brijesh C Purohit

* Working Paper 213/2022
  Potential Inequities in Covid -19 vaccinations
  Brijesh C Purohit

---

* Working papers are downloadable from MSE website http://www.mse.ac.in
$ Restricted circulation