

Effect of Online Searches on Stock Returns

Saurabh Ahluwalia¹, PhD

¹ Assistant Professor, MSC05 3090, Anderson School of Management, 1 University of New Mexico, Albuquerque, NM 87131-0001 USA

Correspondence: Saurabh Ahluwalia, PhD, Assistant Professor, MSC05 3090, Anderson School of Management, 1 University of New Mexico, Albuquerque, NM 87131-0001 USA

Received: October 24, 2017

Accepted: November 13, 2017

Online Published: November 20, 2017

doi:10.5430/afr.v7n1p70

URL: <https://doi.org/10.5430/afr.v7n1p70>

Abstract

I empirically test how online information seeking behavior of investors leads to an impact on asset prices. Various theoretical models assume that information seeking behavior of investors has an impact on asset prices. However, it is very difficult to empirically test this, since the actual information acquisition process of the investors is unobservable. Using a unique data set from Google Trends, I construct a search volume index (SVI) and use it to proxy for the information seeking behavior of retail investors. Using a portfolio based approach I document the asymmetric response of stock returns to the changes in the google search volume index. I find that the portfolio with the highest increase in the SVI has positive and significant alphas while portfolios with decrease in SVI show no significant results. My results are robust to alternative specifications of risk models, time frames and measures of search index. Overall, my results are in line with the hypothesis that retail investors' trades can effect stock prices. The paper's findings add to the debate about information content of the retail trades.

Keywords: Asset Pricing, Google Searches, Retail Investors, Risk Models, Portfolio

1. Introduction

Stock markets act as aggregators of information. Multitudes of investors seek information about different stocks that are traded in the market. It is the information seeking behavior of the investors that makes market prices efficient (Grossman, 1976). What is the mechanism by which the information flows into the market? How does information, dispersed among diverse geographic markets and customer segments, get incorporated into the stock price? Using a dataset based on Google searches, I try to answer these questions.

The process by which information is impounded in stock prices has been studied in detail in the asset pricing literature (Grossman & Stiglitz, 1980; Holthausen & Verrecchia, 1990; Kim & Verrecchia, 1997). An underlying prediction common to these models is that an investor will choose to become informed through information acquisition if the expected benefit of obtaining the information is greater than the cost of acquiring information. However, there has been little empirical research about the actual process used by investors to gather information and trade, since it is hard to observe an investor's information acquisition process. Using search data from Google Trends helps to circumvent this limitation.

I look at the Google searches of company stocks to proxy for the research being done on a company by a large community of investors (Note 1). Google Trends allows users to download search data for the keyword searches done through the Google search engine. Data is in the form of a search volume index that is based on the actual searches done for a particular keyword. I collect data for the S&P 1500 companies over five years from Jan 2004 to Dec 2008. I construct a weekly search volume index (SVI) that represents the volume of searches done for a stock. A high value of SVI indicates that a large group of people are searching for information about a company (Note 2). The search can be motivated by a number of different reasons, such as employment or sales related inquiries. Such queries add noise as they are not investment related. To increase the power of my tests, I restrict attention to searches being done for the ticker symbol of the sample companies. Searches based on the ticker symbol come closest to ascertaining the motive that an individual may have for researching a company.

It is reasonable to assume that primarily individual investors are driving the search volume index since institutional investors will predominantly utilize sophisticated databases such as Bloomberg, and will tend to use outside sources less frequently. Second, searches by institutional investors will be a minuscule percentage of the overall searches done by all investors. Since Google Trends index does not include searches that are below a minimum search volume

threshold, the search volume for keywords searched exclusively by institutional investors will most likely be set to zero. Finally, various studies (Takeda & Wakao, 2014; Da, Engelberg & Gao, 2011) also find that there is a strong and direct correlation between changes in Google searches and trading by retail investors.

I find that change in the SVI has a subsequent impact on weekly and monthly stock returns. Specifically, a portfolio that consists of stocks that experienced the highest increase in month over month SVI earns an annualized risk adjusted excess return of 5.9%. To determine if the impact is permanent or temporary, I look at the returns for up to six months in the future and do not document any significant reversals. Results are similar when weekly changes in SVI are used to form portfolios. These results are consistent with various studies (Kaniel, Saar, & Titman, 2008; Griffin, Shu, & Topaloglu, 2010), that find evidence that retail trading predicts returns around certain corporate events such as earnings and takeover announcements.

I also consider two alternative hypotheses that may explain the results. First, the “liquidity provision” hypothesis states that retail investors provide liquidity to institutional investors by taking the other side of institutional trades. Retail investors demand a premium for providing liquidity to institutional investors and thus exert temporary pressure on prices (Kaniel, Saar & Titman, 2008). In such a case, we should see price reversals in matter of days. Second, it is possible that current period retail buying predicts more retail buying in subsequent periods. The buying pressure on stock leads to high future returns (Dorn, Huberman, & Sengmueller, 2008; Barber, Odean & Zhu, 2009). Such prices increases are also subsequently reversed. However, I do not document any significant reversals for up to six-months.

This paper makes following primary contributions to the literature. I document the actual information seeking process of investors. I analyze the mechanism through which the information seeking behavior leads to price impact in stock market. I find that retail investor trading serves as a conduit for information to flow from disaggregated sources to the market, thus making the market more efficient. My results extend the findings of Da, Engelberg & Gao (2011) who also use data from Google Trends from 2004 to 2008. While Da et al. (2011) show that increase in SVI has a positive effect on stock prices for next two weeks, I show how SVI based trading strategy can generate positive alpha while accounting for the common risk factors, at both monthly and weekly level. Finally, utilizing long term risk adjusted returns I provide evidence supporting information hypothesis (Kaniel et al., 2008) and add to the literature showing that the retail investors can benefit by trading on private information (Griffin et al., 2010).

This paper is organized as follows: section 2 briefly goes over the existing literature; section 3 introduces the hypothesis, Section 4 describes the data and salient issues related with the data used in the paper; section 5 presents the results and section 6 Discusses the results and Section 7 provides limitations and details avenues for further research.

2. Literature Review

2.1 Theoretical Literature

Efficient market hypothesis (Malkiel, 2005) is considered the primary hypothesis that is used as basis to explain how information is reflected in the asset prices. Efficient market hypothesis assumes that asset prices adjust to new information without any delay and there are no arbitrage opportunities possible. Information hypothesis (Kaniel et al., 2008; Griffin et al., 2010; Kelley & Tetlock, 2013) that is based on efficient market hypothesis states that over time information from investors flows into the asset prices through buying and selling of the stocks. On other hand behavioral finance theory suggests that prices do not adjust instantly and nor are they always rational (Barberis & Thaler, 2003). Behavioral finance suggests that public attention alone can influence asset prices. Attention hypothesis (Barber & Odean, 2008) suggests that since individual investors lack the means and time to analyze all of the thousands of stocks available they only make decision to buy stocks based on the limited number of stocks that are part of their limited attention set. While, investor recognition hypothesis (Merton, 1987) states that visibility of stocks encourage investors to buy stocks that they would not have bought otherwise since these now visible stocks were earlier not part of the information set of the investors. Google searches can be considered to provide online visibility to the stocks for the investors (Da et al., 2011).

2.2 Google Trends Literature

Google Trends data has been used in various disciplines and in a variety of settings. Ginsberg et al., (2009) show that one can estimate the level and outbreak of flu activity in different geographical areas in the US by looking at the search terms related to flu. They use the estimate to understand how the flu spreads and how can it be tracked. Choi and Varian (2009) use the Google searches as a proxy for the populations’ demand for information. Choi and Varian (2009) find that the inclusion of the Google searches improves the model for predicting the monthly retail sales of

automobiles and homes. Dorn, Huberman and Sengmueller (2008) and Barber, Odean and Zhu (2009) also find that retail buying predicts short horizon returns. Vlastakis and Markellos (2010) using data based on internet searches find that internet search volume is positively correlated to trading volume and return volatility. Dzielinski (2012) studies how economic uncertainty measured based on google searches can effect stock returns. Dimpfl and Kleiman (2016) model investor sentiment using google searches and show how it is negatively related to stocks in Germany.

This paper is also related to Da, Engelberg and Gao (2011) who use data from Google Trends for a sample of Russell 3000 stocks from 2004 to 2008. They find that the Google SVI predicts future retail volume and an increase in the SVI contributes to large first-day returns and subsequent long run under-performance of IPOs. Their results provide support for Barber and Odean (2007, 2008) naïve investors' buying pressure on the price hypothesis. They find some evidence that the SVI predicts higher stock prices over the next two weeks. However, the price increase is subsequently reversed.

Current paper is close to Kaniel, Saar and Titman (2008) who look at short horizon returns subsequent to net buying by individual investors for 1,920 NYSE stocks from 2000 through 2003. They find stocks that are heavily bought by individuals in one week consistently outperform the market the following week. Kaniel et al. (2008) propose liquidity provision hypothesis that states that risk-averse individual investors provide liquidity to institutions that demand immediacy. Hence, they suggest that as institutions sell to individuals prices fall one week and rebound the next. While I also look at the impact of retail investors on stock prices, I differ from Kaniel et al. (2008) in a number of key respects. I use a more primitive measure of retail investors' propensity to search for information as represented by the changes in the Google SVI, to determine the impact of retail investors on the stock prices. I also differ from Kaniel et al. (2008) in terms of data and time frame under consideration.

3. Hypothesis

An increase in google searches for a firm can indicate that retail investors are indulging in information-seeking behavior and are actively seeking information about the company. This behavior could have been triggered by some private information, news or some other observation about the firm. If the investors' research sheds a positive light on the company, they end up buying it. However, if the investors uncover something negative, they tend not to do anything since retail investors tend not to short-sell and because of short selling constraints and costs (Diamond & Verrecchia, 1987; Da et al., 2011). Figure 1. represents conceptually the flow from information seeking behavior to asset prices.

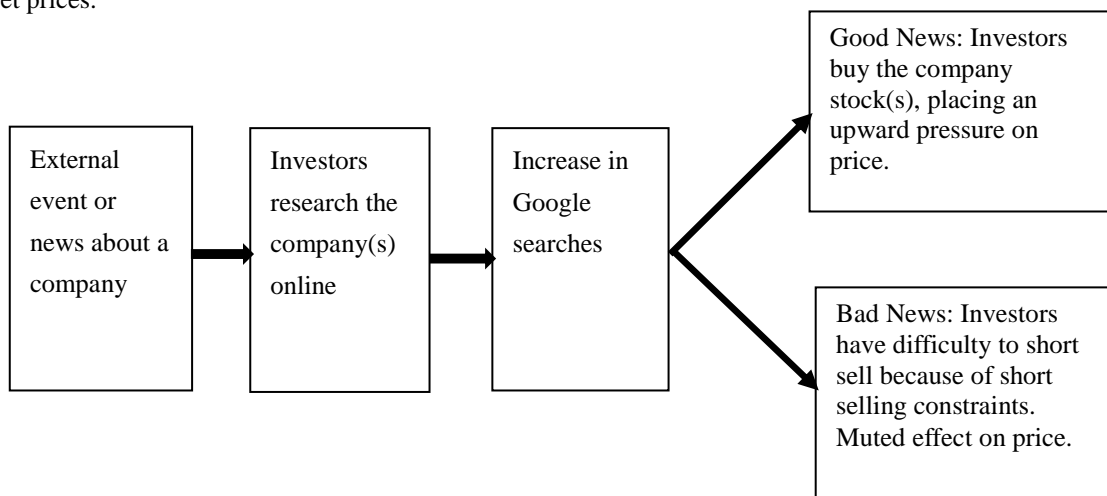


Figure 1. Information flow to asset prices

On other hand (Barber & Odean, 2008; Merton,1987) naïve investor attention hypothesis posit that a mere increase in attention by uninformed, under diversified retail investors cannot lead to any permanent increase in stock price and a decrease in attention can lead to an increase in future prices. Hence, we cannot know for certain based on the theory how the changes in SVI will impact the future returns. We examine the question empirically by stating our hypothesis as follows:

Hypothesis 1:

H_0 : Increase in SVI is not associated with increase in future stock returns.

H_1 : Increase in SVI is associated with increase in future stock returns

Hypothesis 2:

H₀: Decrease in SVI is not associated with decrease in future stock returns.

H₁: Decrease in SVI is associated with decrease in future stock returns.

We want to ascertain if the price impact is permanent or not. One possibility, consistent with the “information hypothesis or permanent price impact,” is that the increase in the SVI leads to information flow into the stock prices. It is also possible that the stock returns increased because the retail investors were providing short-term liquidity to the institutional investors (Kaniel et al., 2008) . One would expect price reversals in a matter of days once the short-term liquidity needs of the institutional investors are satisfied. Thus we state our third hypothesis:

Hypothesis 3:

H₀: Price increase experienced due to changes in SVI is not reversed in subsequent time periods.

H₁: Price increase experienced due to changes in SVI is reversed in subsequent time periods.

4. Data and Methodology

Google Trends allows users to download data about volume of keyword searches done through the Google search engine. Data is available from Jan 2004 onwards and was first made available to the public in Jan 2008. Google Trends data does not report the raw number of searches that were done for a particular keyword, but does report a search volume index, calculated via the following two step procedure: First, the query share is calculated by dividing the total query volume for a search keyword by the volume of the total number of queries for all the search keywords, and, second, the query share index is normalized so that it starts at 0 in January 1, 2004. Hence, the data at later dates is the percentage deviation from the query share on January 2004.

SVI data is available at the state level for the United States as Google is able to determine the geographical area of the user by decoding the IP address of the user making the query. Hence, using Google Trends one can download the SVI that represents the search volume for a particular keyword for a given region at a particular point in time. I download the weekly (US only) SVI for S&P 1500 Composite Index firms from Jan 2004 to Dec 2008. S&P Composite 1500 firms account for almost 90% of the U.S. market capitalization. I restrict the sample to S&P 1500 list of companies to make the task of data collection manageable.

The ticker symbol of each of the S&P 1500 firms was chosen as the keyword for which the SVI data was downloaded. There are several advantages to using the ticker symbol. An individual may search for a company for a variety of different reasons such as looking for a job, checking a company product or some other non-investment related purpose. Such searches are devoid of any informational content for the stock price. Ideally, I would like to consider only those searches that are related to an investment purpose. At present, it is impossible to do so, since Google does not provide any contextual information for keyword searches. Hence, as a proxy for searches done for investment-related reasons, I look at the ticker symbol searches for the companies.

There are a few ticker symbols that stand for common words such as “WIN,” “CAT,” or “TRY.” Hence, I exclude 178 tickers, from my sample, that also stand for common words. The SVI is biased towards larger companies because Google Trends data truncates the searches if they are below a certain threshold. Tickers that do not get searched often or have search frequency so low that it does not meet Google Trends’ minimum search volume requirement have their SVI set at zero. Out of the 388,321 firm-week observations 196,837 are zeroes. Observations with the SVI equal to zero are concentrated in predominantly small stocks. It is only natural that the small stocks that are less well-known and have fewer shareholders and employees will tend to be searched less often. I exclude the stocks for which the SVI is zero for more than 90% of the weeks over the five-year period. This results in the exclusion of 391 stocks.

Data provided by Google Trends has a seasonal and growth component. From 2004 to 2008, both the market share of the Google search engine and the number of people accessing the internet has increased. To account for such mechanical properties of the data, I divide the weekly SVI for each stock by an index composed of the average SVI of the three most common investment related keywords (Google finance, Yahoo finance, and DJIA) searched by users. To ensure that my results are robust to the choice of the above keywords, I also used an index composed of the 10 most common keywords, with similar results.

I employ portfolio based approach utilizing OLS regressions to analyze the question how changes in SVI impact future stock returns. Alpha based approach is widely used in finance literature to measure performance of a portfolio and is considered a central measure to evaluate performance of a trading strategy (Fama & French, 1992, 1993; Ferson & Lin, 2014; Levy & Roll, 2016). By calculating risk adjusted returns for the portfolios we can determine the

alpha that can help to measure the effectiveness of trading strategy based on the changes in SVI. Portfolio based approach has several advantages to using regressions based on individual companies such as lack of endogeneity problems, less noisy and possibility of having a trading strategy that can give excess returns while accounting for common risk factors. It is debatable in how much time will a change in SVI effect stock returns. Hence, I use monthly and weekly time frames to test if the effect is more short term or long term. Further I look at returns up to six months in future to test if the prices show any reversals.

5. Empirical Results

5.1 Summary Statistics

Table 1 shows the distribution of the S&P 1500 companies by industry for the year (2006). Manufacturing has the most representation in the sample while Food Products has the least representation.

Table 1. Summary Statistics of the sample companies based on industry.

SIC Codes	Industry Description	Number Of Firms	%
0-1000	Food Products	7	0.47%
1000-1999	Mining and Construction	78	5.20%
2000-2999	Consumer Products	216	14.40%
3000-3999	Manufacturing	375	25.00%
4000-4999	Utilities and Transportation	148	9.87%
5000-5999	Wholesale and Retail	172	11.47%
6000-6999	Financial Services	268	17.87%
7000-7999	Personal & Business Services	159	10.60%
8000-8999	Miscellaneous	77	5.13%
Total		1500	100.00%

Description: Above table shows the industry distribution of S&P 1500 firms. The sample period is from January 2004 to December 2008.

Table 2 presents salient statistics for the sample firms. The S&P 1500 index includes 500 large firms in Standard & Poor's 500 (S&P 500), 400 midcap firms in the Standard & Poor's Midcap 400 (S&P 400) and 600 small firms in the Standard & Poor's SmallCap 600 (S&P 600). Hence, we get a very representative sample of the firms when we utilize S&P 1500 as shown in Table 2.

Table 2. Summary statistics for the sample S&P 1500 firms.

	Mean	Std. Dev.	Min	Max	NumObs
Assets	18,417	108,087	4.07	1,668,498	1500
MktValue	9,115	25,544	21.10	368,223	1494
LnSale	5.812	1.781	-3.863	11.286	1500
Cashflowat	0.025	0.030	-0.313	0.280	1348
RetonAssets	0.015	0.027	-0.314	0.280	1494
MTB	2.235	1.642	0.582	16.238	1494
LevBook	0.294	0.244	0.000	1.011	1383
LevMkt	0.169	0.178	0.000	0.895	1373

Description: Variables used in the above table are calculated as follows: Assets is total assets (Compustat item atq). MktValue is market value of common equity and is defined as common shares outstanding times the quarter end price (Compustat cshoq*prccq). LnSale is natural log of sales (Compustat item saleq). Cashflowat is the sum of income before extraordinary items and depreciation and amortization divided by lagged assets value (Compustat (ibq+dpq)/l.atq). RetonAssets is income before extraordinary items divided by lagged assets value (Compustat ibq/l.atq). MTB is market to book ratio and is defined as the sum of total assets and market value of equity minus the book value of equity divided by total assets (Compustat (atq-seqq+cshoq*prccq)/atq). LevBook is defined as long-term debt plus current liabilities divided by long-term debt plus current liabilities plus book value of common equity (Compustat (dlcq + dltdq) / (dlcq + dltdq + ceqq)). LevMkt is defined as long-term debt plus current liabilities divided by long-term debt plus current liabilities plus market value of common equity (Compustat (dlcq + dltdq) / (dlcq + dltdq + cshoq*prccq)). The sample period is from January 2004 to December 2008.

Table 3 shows the summary statistics for the SVI, averaged monthly, for each year. There are clear outliers in the data (exceeding the mean by over 20 standard deviations) and the results in this table and the subsequent tables are winsorized at 1% and 99% level. Also shown is the average of the month over month differences. The mean change is slightly positive for each year and varies between -3.2 to 3.5.

Table 3. Summary statistics of the monthly search index for the sample S&P 1500 firms.

Monthly Search Index					
Year	Obs.	Mean	Std. Dev.	Min	Max
2004	9946	0.580	0.513	0.000	3.063
2005	9923	0.710	0.499	0.000	3.575
2006	9897	0.848	0.463	0.000	3.480
2007	9884	1.010	0.492	0.000	3.538
2008	9802	1.155	0.582	0.000	3.575

Month Over Month Change in Search Index					
Year	Obs.	Mean	Std. Dev.	Min	Max
2004	9110	0.008	0.180	-2.650	2.930
2005	9911	0.012	0.205	-2.650	2.650
2006	9880	0.010	0.226	-2.800	3.350
2007	9868	0.017	0.270	-3.200	3.525
2008	9760	0.003	0.214	-2.108	3.050

Description: The monthly search index is the average of the weekly search index for the month. The sample period is from January 2004 to December 2008.

5.2 Weekly Portfolio Rebalancing

In the following section, I analyze the performance of the portfolios formed based on the changes in the SVI. I want to test; 1) If the portfolio with greater positive change in the prior SVI earns higher excess returns. 2) If the performance of the portfolio can be explained by the style tilts or factor loadings. 3) If there are any reversals in the future time periods. Since my sample period is small (five years), I look at the weekly returns of the portfolios formed on bases of change in the SVI. I define change in the SVI as the difference in the SVI for week 5 minus the average SVI for week 1 to week 4. Each week, I divide the stocks into five portfolios based on the changes in the SVI. Hence, portfolio 5 at end of week w will have stocks that had experienced the most increase in the SVI from the prior four week average SVI. Similarly, portfolio 1 at end of week w will have stocks that had experienced the most decrease in the SVI from the prior four week average SVI. Every week the portfolios are rebalanced and stocks are assigned to new quintiles. Excess returns are calculated for week $w+1$ by subtracting the risk-free rate. Excess returns are regressed against the risk factors, as specified by equation (1) below:

$$R_{i,t}^e = \alpha_{i,t} + \beta_i R_{m,t}^e + s_i SML_t + h_i HML_t + \varepsilon_{i,t} \quad (1)$$

$$i = 1, 2, \dots, 5$$

$$t = 1, 2, \dots, T$$

Here $R_{i,t}^e$ is the value-weighted return of the portfolio i in week t in excess of the benchmark risk free rate. Alpha is the primary variable of interest and it shows if the trading strategy is generating risk adjusted excess returns. $R_{i,t}^e$, SML , HML , MOM are respectively the excess return on the market, size and value factors taken from Ken French's website. To account for serial correlation, I use the Newey West correction for standard errors.

Table 4 shows the results of regressing the weekly excess returns on the Fama-French (1993) (Note 3) risk factors. This table shows the results of weekly regressions of the excess portfolio returns on the Fama-French(1993) factors, $R_m - R_f$, HML and SMB . Excess returns are calculated by subtracting the risk free rate from the value-weighted portfolio returns. $R_m - R_f$ is the value-weighted market return on all NYSE/Amex/Nasdaq firms minus the risk-free

rate. SMB (small minus big) is the difference each week between the return on small- and large-capitalization firms. HML (high minus low) is the difference each week between the return of a portfolio of high book-to-market stocks and the return of a portfolio of low book-to-market stocks. Each week the stocks in the S&P 1500 index are assigned to one of the five portfolios based on the weekly change in the search index. Stocks experiencing the largest week over week increase are assigned to portfolio five, while the stocks that experience the largest week over week decrease are assigned to portfolio 1. Specifications 7 to 12 in panel B show results for portfolio 5 for subsequent weeks. For example, p5 W+2 constitutes one week returns, two weeks after the date of portfolio formation.

The results show that if one were to buy the stocks in the highest quintile one week after the increase in the index, one can earn a weekly alpha of 0.132%, significant at 1% level. The 132 basis point weekly spread translates into an annualized return of 6.86%. In light of the average annualized return for S&P 500 of about 11% the annualized return of 6.86% is economically significant. Column 6 shows that the trading strategy of buying the stocks in portfolio 5 and shorting the stocks in portfolio 1 does not yield a significant alpha. The above results show that an increase in the SVI is associated with a subsequent positive price impact on stocks supporting hypothesis 1 and 2.

Table 4 also shows the results for holding the portfolio at future weeks up to two months after the portfolio formation date. I do not document any reversals expected to take place in the future. An absence of reversals indicates that the price impact is not short-lived, as will be predicted by the “liquidity provision” hypothesis. If retail investors were simply providing short-term liquidity to the institutional investors, one would expect price reversals in a matter of days (Kaniel et al., 2008).

Table 4. Weekly risk-adjusted excess returns of the portfolios based on SVI changes.

Panel A						
	potrfolio 1 (p1) - largest decrease in SVI	potrfolio 2	potrfolio 3	potrfolio 4	potrfolio 5 (p5) - largest increase in SVI	p5-p1
	(1)	(2)	(3)	(4)	(5)	(6)
Alpha	0.00035 (0.0005)	0.00043 (0.0005)	0.00032 (0.0007)	0.00008 (0.0006)	0.00132*** (0.0005)	0.00097 (0.0006)
Rm-Rf	0.855*** (0.0487)	0.878*** (0.0314)	0.794*** (0.0784)	0.868*** (0.0345)	0.885*** (0.0317)	0.030 (0.0371)
SMB	0.030 (0.0778)	0.009 (0.0542)	0.251 (0.1265)	-0.081 (0.0628)	-0.001 (0.0458)	-0.030 (0.0705)
HML	-0.195*** (0.0490)	-0.213*** (0.0592)	-0.029 (0.0789)	-0.163** (0.0711)	-0.139** (0.0589)	0.055 (0.0629)
N	292	292	292	292	292	292
R Sq.	0.91	0.91	0.83	0.85	0.89	0.02

Panel B

	p5 (w+2 week)	p5 (w+3 week)	p5 (w+4 week)	p5 (w+8 week)	p5 (w+9 week)	p5 (w+10 week)
	(7)	(8)	(9)	(10)	(11)	(12)
Alpha	-0.00030 (0.0005)	0.00074 (0.0005)	0.00044 (0.0004)	0.00020 (0.0005)	0.00076 (0.0005)	0.00038 (0.0005)
Rm-Rf	0.878*** (0.0442)	0.872*** (0.0249)	0.883*** (0.0278)	0.813*** (0.0587)	0.838*** (0.0540)	0.926*** (0.0341)
SMB	0.038 (0.0760)	-0.058 (0.0460)	-0.064 (0.0433)	0.059 (0.0932)	0.053 (0.0889)	-0.004 (0.0603)
HML	-0.162*** (0.0487)	-0.155** (0.0512)	-0.114** (0.0525)	-0.071 (0.0454)	-0.061 (0.0483)	-0.156** (0.0554)
N	292	292	292	289	288	287
R Sq.	0.91	0.89	0.91	0.88	0.89	0.91

Notes:

1. Robust standard errors are in parentheses.
2. The sample period is from January 2004 to December 2008.
3. ***, **, * denote 1%, 5% and 10% significance levels, respectively.

5.3 Monthly Portfolio Rebalancing

As a further robustness check, I rerun the above analysis on a monthly basis. To test whether an increase in search activity predicts a future price increase that leads to incorporation of information in stock prices and subsequent price rise, I form five portfolios based on the month over month change in the average search index. Specifically, I divide the stocks into five portfolios based on changes in the SVI from one month to the next. Hence, portfolio 5 at the end of month *m* will have stocks that have experienced the most increase in the SVI from month *m-1* to month *m*. Similarly, portfolio 1 at the end of month *m* will have stocks that have experienced the most decrease in the SVI from month *m-1* to month *m*. Every month the portfolios are rebalanced and stocks are assigned to new quintiles based on the change in the index from previous month. Excess returns are calculated for month *m+1* by subtracting the risk free rate. Excess returns are regressed against the risk factors, as specified by equation (1) below:

$$R_{i,t}^e = \alpha_{i,t} + \beta_i R_{m,t}^e + s_i SML_t + h_i HML_t + m_i MOM_t + \varepsilon_{i,t} \quad (2)$$

$$i = 1, 2, \dots, 5$$

$$t = 1, 2, \dots, T$$

Here $R_{i,t}^e$ is the return of the portfolio *i* in month *t* in excess of benchmark risk free rate.

α is the primary variable of interest and it shows if the trading strategy is generating risk adjusted excess returns.

R^e , SML , HML , MOM are respectively the excess return on the market, size, value and momentum factors taken from Ken French's website (Note 4). To account for serial correlation, I use the Newey West correction for standard errors.

Table 5 shows the results of the monthly regressions. This table shows the results of monthly regressions of portfolio returns on the four Cahart(1997) factors, Rm-Rf, HML, SMB and Mom (momentum). Dependent variable is the value-weighted return less the risk free rate. Rm-Rf is the value-weighted market return on all NYSE/Amex/Nasdaq firms minus the risk-free rate. SMB (small minus big) is the difference each month between the return on small- and large-capitalization firms. HML (high minus low) is the difference each month between the return of a portfolio of high book-to-market stocks and the return of a portfolio of low book-to-market stocks. Each month the stocks in the S&P 1500 index are assigned to one of the five portfolios based on the month over month change in the search index. P5 is the portfolio that includes stocks that have had the largest increase in the search index from the previous month. Similarly, P1 is the portfolio that includes stocks that have had the largest decrease in the search index from the

previous month. Specifications 7 to 11 show results for portfolio 5 if held for additional months. For example, p5 (m+2) looks at monthly returns two months in future.

Table 5. Monthly risk-adjusted excess returns of the portfolios based on SVI changes.

Panel A						
	potrfolio 1 (p1) - largest decrease in SVI	portfolio 2	portfolio 3	portfolio 4	potrfolio 5 (p5) - largest increase in SVI	p5-p1
	(1)	(2)	(3)	(4)	(5)	(6)
Alpha	0.0029 (0.0021)	-0.0007 (0.0017)	0.0001 0.0023	0.00022 (0.0016)	0.0049** (0.0019)	0.002 (0.0030)
Rm-Rf	0.88*** (0.0640)	0.83*** (0.0520)	0.95*** 0.06	0.82*** (0.0420)	0.77*** (0.0530)	-0.11 (0.0800)
SMB	-0.12 (0.1340)	-0.0033 (0.1070)	-0.236* 0.124	-0.018 (0.0980)	0.103 (0.0973)	0.22 (0.1800)
HML	-0.13 (0.0850)	0.013 (0.0760)	0.034 0.114	0.266*** (0.0790)	-0.12 (0.0720)	0.011 (0.1180)
Momentum	-0.059 (0.1000)	-0.062 (0.0624)	-0.179*** 0.06	-0.119* (0.0576)	-0.093 (0.0590)	-0.034 (0.1490)
N	56	56	56	56	56	56
R Sq.	0.86	0.9	0.88	0.93	0.9	0.05
Panel B						
	p5 (m+2 month)	p5 (m+3 month)	p5 (m+4 month)	p5 (m+5 month)	p5 (m+6 month)	p5 upto Dec 2007
	(7)	(8)	(9)	(10)	(11)	(12)
Alpha	0.003439* (0.0020)	-0.001757 (0.0019)	-0.003104 (0.0020)	0.0006833 (0.0023)	0.0001818 (0.0018)	0.0053** (0.0022)
Rm-Rf	0.8377*** (0.0864)	0.8556*** (0.0560)	0.7919*** (0.0599)	0.8742*** (0.0798)	0.7956*** (0.0479)	0.84*** (0.0912)
SMB	-0.1921 (0.1278)	0.04379 (0.1163)	0.03651 (0.0995)	-0.03306 (0.1302)	-0.05414 (0.1053)	0.0964 (0.1014)
HML	0.01371 (0.1244)	0.04176 (0.1022)	-0.1034 (0.0972)	-0.07459 (0.1177)	0.09220 (0.0697)	-0.193* (0.1103)
Momentum	-0.08947 (0.0610)	-0.02224 (0.0375)	-0.08002 (0.0483)	0.09191 (0.0638)	-0.01891 (0.0332)	-0.185** (0.0743)
N	56	56	56	56	56	44
R Sq.	0.89	0.86	0.87	0.84	0.82	0.78

Notes:

1. Robust standard errors are in parentheses.
2. The sample period is from January 2004 to December 2008.
3. ***, **, * denote 1%, 5% and 10% significance levels, respectively.

Alpha for portfolio 5, the highest positive SVI change portfolio, is significant at a 10% level. Based on one month returns, the alpha translates into an annualized excess return of about 5.9%. The weekly alpha is about one percent higher than the alpha for monthly returns. However, in case of weekly portfolio rebalancing, the effective returns are diminished by the transaction costs accruing because of the increased trading frequency. I also look at the performance of the portfolio up to six months in advance and I do not document any significant price reversals.

6. Discussion

The results show that the search index-representative of volume of searches being done for a particular stock- has a permanent price impact. I find that the portfolio with the largest increase in the search index has a significant risk adjusted alpha that translates into an annualized return of 5.9%. The results lead us to reject null hypothesis 1 and provide support for the alternative hypothesis. Comparing the portfolios with increase vs decrease in SVI shows that stock prices have an asymmetric response to the changes in SVI. While increase in SVI leads to an increase in future returns the reverse is not true. Hence, we accept null hypothesis 2 that a decrease in SVI does not lead to a decrease in future stock returns. It further supports that the increase in SVI is not merely attention as mentioned in Da et al., (2011) but is also related to information seeking or research related behavior of the investors. In case of SVI representing merely attention a decrease in SVI will represent a decrease in attention and in that case both Barber and Odean's (2008) attention hypothesis and Merton's 1987 investor recognition hypothesis will imply that a decrease in attention should result in higher returns.

The results lead us to accept the null hypothesis 3 since the price impact seems permanent and I do not document any reversals in the price over subsequent periods of up to six months. The results are similar when portfolios are rebalanced each week based on the week over week changes in the search index. I find support for the information hypothesis of Kaniel et al., (2008), that retail investor information seeking behavior as captured by search index has information content that has a permanent price impact. The results are also consistent with Griffin et al., (2010) and Kaniel, Liu, Saar and Titman, (2012) who also find that retail investors can trade for informational reasons. It is natural to ask why the market is not able to discern the effect of the increase in the SVI on the returns. One of the reasons is that the information on aggregated searches has been public only since Jan 2008 and the market had no way of accessing this information prior to this period.

7. Limitations and Further Research

Our study is subject to several limitations based on the source of data from Google Trends. Google Trends data is not available for very low volume keyword searches, hence, less well-known companies can be actively searched but if the number of searches does not meet the minimum threshold then Google does not provide any information on SVI for these companies. Since many small companies can fall into this category we expect that our results are not broadly applicable to small and less well known firms. We hope Google provides more micro level data in future and study such as ours can be replicated for small company stocks too. Further, if the search data is available in context we can further sharpen our analysis. For example, in our study we assume that the search for ticker symbols is related to investor seeking information about the company. We can be more certain about our conclusions if Google can provide some data that categorizes user searches based on the intent of the user. Asset pricing models used in the paper, while standard and widely used, assume that the risk of the portfolio can be accounted by three or four factors. In future, if a better asset pricing model is discovered the results should be run based on the new model.

8. Conclusion

Using online searches for a ticker symbol on the internet as a proxy for investors' information seeking behavior, I look at the impact of the searches on stock prices. Data from Google trends allows me to proxy for the active interest of millions of investors in a stock. It is possible that some event or information may have triggered the interest which leads investors to further research a company. Since the internet is the biggest news and information source available to a retail investor, it is expected that one will seek stock information online. Moreover, I observe that investor response is asymmetrical for price increases vs. decreases since individual investors have short selling constraints. Results show that portfolios formed on basis of increase in SVI yield economically significant annualized returns ranging from 5.9% to 6.9%. The paper provides evidence that information seeking behavior of retail investors can be captured by google searches and that subsequent increase in SVI has an effect on asset prices.

References

- Barber, B. M., & Odean, T. (2007). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The Review of Financial Studies*, 21(2), 785-818. <https://doi.org/10.1093/rfs/hhm079>
- Barber, B. M., Odean, T., & Zhu, N. (2008). Do retail trades move markets?. *The Review of Financial Studies*, 22(1), 151-186. <https://doi.org/10.1093/rfs/hhn035>
- Barber, B. M., Odean, T., & Zhu, N. (2009). Systematic noise. *Journal of Financial Markets*, 12(4), 547-569. <https://doi.org/10.1016/j.finmar.2009.03.003>

- Barberis, N., & Thaler, R. (2003). A survey of behavioral finance. *Handbook of the Economics of Finance, 1*, 1053-1128. [https://doi.org/10.1016/S1574-0102\(03\)01027-6](https://doi.org/10.1016/S1574-0102(03)01027-6)
- Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance, 52*(1), 57-82. <https://doi.org/10.1111/j.1540-6261.1997.tb03808.x>
- Choi, H., & Varian, H. (2009). Predicting initial claims for unemployment benefits. *Google Inc*, 1-5.
- Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *The Journal of Finance, 66*(5), 1461-1499. <https://doi.org/10.1111/j.1540-6261.2011.01679.x>
- Diamond, D. W., & Verrecchia, R. E. (1981). Information aggregation in a noisy rational expectations economy. *Journal of Financial Economics, 9*(3), 221-235. [https://doi.org/10.1016/0304-405X\(81\)90026-X](https://doi.org/10.1016/0304-405X(81)90026-X)
- Diamond, D. W., & Verrecchia, R. E. (1987). Constraints on short-selling and asset price adjustment to private information. *Journal of Financial Economics, 18*(2), 277-311. [https://doi.org/10.1016/0304-405X\(87\)90042-0](https://doi.org/10.1016/0304-405X(87)90042-0)
- Dimpfl, T., & Kleiman, V. (2016). Investor Pessimism and the German Stock Market: Exploring Google Search Queries. *German Economic Review*. <https://doi.org/10.2139/ssrn.2801698>
- Dorn, D., Huberman, G., & Sengmueller, P. (2008). Correlated trading and returns. *The Journal of Finance, 63*(2), 885-920. <https://doi.org/10.1111/j.1540-6261.2008.01334.x>
- Drake, M. S., Roulstone, D. T., & Thornock, J. R. (2012). Investor information demand: Evidence from Google searches around earnings announcements. *Journal of Accounting Research, 50*(4), 1001-1040. <https://doi.org/10.1111/j.1475-679X.2012.00443.x>
- Dzielinski, M. (2012). Measuring economic uncertainty and its impact on the stock market. *Finance Research Letters, 9*(3), 167-175. <https://doi.org/10.1016/j.frl.2011.10.003>
- Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *The Journal of Finance, 47*(2), 427-465. <https://doi.org/10.1111/j.1540-6261.1992.tb04398.x>
- Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics, 33*(1), 3-56. [https://doi.org/10.1016/0304-405X\(93\)90023-5](https://doi.org/10.1016/0304-405X(93)90023-5)
- Ferson, W., & Lin, J. (2014). Alpha and performance measurement: the effects of investor disagreement and heterogeneity. *The Journal of Finance, 69*(4), 1565-1596. <https://doi.org/10.1111/jofi.12165>
- Francis, J., K. Schipper, & L. Vincent. (2002). Expanded Disclosures and the Increased Usefulness of Earnings Announcements. *The Accounting Review, 77*(3), 515-546. <https://doi.org/10.2308/accr.2002.77.3.515>
- Ginsberg, J., Matthew H. M., Rajan S. P., Lynnette B., Mark S. S., & Larry B., 2009, Detecting influenza epidemics using search engine query data, *Nature, 457*, 1012-1014. <https://doi.org/10.1038/nature07634>
- Griffin, J. M., Shu, T., & Topaloglu, S. (2010). Examining the dark side of financial markets: Who trades ahead of major announcements. In *SSRN: 1363756 Goetzmann WN, Ingersoll JE, Ross SA, 2003. High-Water Marks and Hedge Fund Management Contracts. The Journal of Finance, 58*. <https://doi.org/10.1093/rfs/hhs058>
- Grossman, S. (1976). On the efficiency of competitive stock markets where trades have diverse information. *The Journal of Finance, 31*(2), 573-585. <https://doi.org/10.1111/j.1540-6261.1976.tb01907.x>
- Grossman, S. J., & Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *The American Economic Review, 70*(3), 393-408.
- Harris, M., & Raviv, A. (1993). Differences of opinion make a horse race. *The Review of Financial Studies, 6*(3), 473-506. <https://doi.org/10.1093/rfs/5.3.473>
- Holthausen, R. W., & Verrecchia, R. E. (1990). The effect of informedness and consensus on price and volume behavior. *Accounting Review, 191*-208.
- Imhoff, E.A. 1992. The Relation between Perceived Accounting Quality and Economic Characteristics of the Firm, *Journal of Accounting and Public Policy, 11*, 97-118. [https://doi.org/10.1016/0278-4254\(92\)90019-T](https://doi.org/10.1016/0278-4254(92)90019-T)
- Narasimhan, J. 1990. Evidence of predictable behavior of security returns, *Journal of Finance, 45*, 881-898. <https://doi.org/10.1111/j.1540-6261.1990.tb05110.x>
- Kaniel, R., Saar, G., & Titman, S. (2008). Individual investor trading and stock returns. *The Journal of Finance, 63*(1), 273-310. <https://doi.org/10.1111/j.1540-6261.2008.01316.x>

- Kaniel, R., Liu, S., Saar, G., & Titman, S. (2012). Individual investor trading and return patterns around earnings announcements. *The Journal of Finance*, 67(2), 639-680. <https://doi.org/10.1111/j.1540-6261.2012.01727.x>
- Kelley, E. K., & Tetlock, P. C. (2013). How wise are crowds? Insights from retail orders and stock returns. *The Journal of Finance*, 68(3), 1229-1265. <https://doi.org/10.1111/jofi.12028>
- Kim, O., & Verrecchia, R. E. (1997). Pre-announcement and event-period private information. *Journal of Accounting and Economics*, 24(3), 395-419. [https://doi.org/10.1016/S0165-4101\(98\)00013-5](https://doi.org/10.1016/S0165-4101(98)00013-5)
- Lehmann, B. N. (1990). Fads, martingales, and market efficiency. *Quarterly Journal of Economics*, 105, 1-28. <https://doi.org/10.2307/2937816>
- Levy, M., & Roll, R. (2016). Seeking Alpha? It's a Bad Guideline for Portfolio Optimization. *The Journal of Portfolio Management*, 42(5), 107-112. <https://doi.org/10.3905/jpm.2016.42.5.107>
- Malkiel, B. G. (2005). Reflections on the efficient market hypothesis: 30 years later. *Financial Review*, 40(1), 1-9. <https://doi.org/10.1111/j.0732-8516.2005.00090.x>
- Merton, R. C. (1987). A simple model of capital market equilibrium with incomplete information. *The Journal of Finance*, 42(3), 483-510. <https://doi.org/10.1111/j.1540-6261.1987.tb04565.x>
- Schwert, G. W. (1981). The adjustment of stock prices to information about inflation. *The Journal of Finance*, 36(1), 15-29. <https://doi.org/10.1111/j.1540-6261.1981.tb03531.x>
- Takeda, F., & Wakao, T. (2014). Google search intensity and its relationship with returns and trading volume of Japanese stocks. *Pacific-Basin Finance Journal*, 27, 1-18. <https://doi.org/10.1016/j.pacfin.2014.01.003>

Notes

Note 1. According to the Search Engine Journal, Google search engine had 79.88% market share. Hence, Google searches for a topic can be considered representative of overall searches being done for a topic.

Note 2. There are many different ways investors can research a stock, such as reading analyst reports, reading newspapers, etc. However, in the past decade the internet has become the primary source of news and information in US. In a recent survey conducted in Jan 2011 by the Pew Project for Excellence in Journalism, (<http://www.journalism.org>) 34% of respondents said they read news online within the past 24 hours (in contrast to 31% who favored newspapers), and a full 41% said they get most of their news online (10% more than those who said they got most of their news from a newspaper).

Note 3. We cannot apply momentum factor in weekly analysis since weekly returns show short term reversals (Jegadeesh, 1990 and Lehmann 1990)

Note 4. Ken French maintains an extensive database of Fama-French Factors and can be freely accessed through his website- http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html