Shuheng Dai

Shanna Su

Jin-Myung Yoo

Rudolf Pham

Sang Lee

Google Trends vs. Stock Performance

# Final Project

## Stat 133

**Introduction**

Financial trading data sets reflect the myriad of decisions taken by market participants. Investors begin their decision-making processes by attempting to gather information. Investors’ sentiments may be reflected in search results of key words in timely bases. In today's world, information gathering often consists of searching online sources. Recently, the search engine Google has begun to provide access to aggregated information on the volume of queries for different search terms and how these volumes change over time, via the publicly available service Google Trends. In this project, we investigate the intriguing possibility of analyzing search query data from Google Trends to provide new insights into the information gathering process that precedes the trading decisions recorded in the stock market data. By analyzing changes in Google query volumes for search terms related to each company, we would like to find patterns that may be interpreted as “early warning signs” of stock market moves.

**Method**

**1) Data**

Our goal is to explore longitudinal trends in the stock market using keyword frequencies as our variables and to discover differences among the companies. We will be looking at three specific companies from each of four different industries listed on the U.S. stock market using a tentative time range of roughly three years: from January 1, 2011 to November 30, 2013.

We first identified a set of 60 significant search terms, five for each company. We web-scraped company profiles from well-known media such as “Wall Street Journal”, “CNN”, “Reuters”, “NASDAQ”, “Business Week”, and “Yahoo”, and using R, we were able to identify semantically related keywords and ranked them in order of frequent usage. We then grouped some of the similar words together such as “Advertisement” and “Ads”, in order to reduce the repetitive data mining. Through this process, we intentionally introduced some financial bias, taking only the relevant, frequently used words related to each company.

|  |  |
| --- | --- |
| Companies | Key Words |
| **Hot Tech Firms** |  |
| Apple (APPL) | Apple, ios, app, itunes, mac |
| Google (GOOG) | ads, Google, Chrome, search, YouTube |
| Facebook (FB) | Facebook, developer, share, social, user |
| **Specialty Eateries** |  |
| Starbucks (SBUX) | beverage, coffee, Starbucks, Tazo, Teavana |
| Panera Bread (PNRA) | bread, dough, Panera, bake, bakery |
| Potbelly (PBPB) | breakfast, chocolate, Potbelly, salad, sandwich |
| **Auto Manufacturers** |  |
| Tesla Motors (TSLA) | battery, powertrain, roadster, Tesla, vehicle |
| General Motors (GM) | automotive, dealer, Ford, motor, vehicle |
| Ford Motors (F) | automotive, Chevrolet, GM, motor, vehicle |
| **Fashion** |  |
| The Gap (GPS) | apparel, Banana Republic, Gap, old navy, Piperlime |
| Coach (COH) | accessories, bags, Coach, footwear, fragrance |
| Steve Madden (SHOO) | accessories, fashion, footwear, retailer, Steve Madden |

We used Google Trends to mine data on the frequency of search terms over the three year period. The search counts were normalized so that the average frequencies over the time period for each search term were 100. We then mined stock price data from the corresponding period using Yahoo finance, compiled and analyzed them all together, then calculated the strength of correlation between all 60 keywords and their corresponding company’s stock performances. Next, we analyzed the newly formed data sets and created time intervals of 30 days to compare stock prices with search volumes, then repeated the same process with time intervals of 30 days. Our goal was to analyze the data on keyword frequencies and stock market performances for different companies, identify the relation between stock performance and search volume, observe changes in correlations, and to observe any relations between stock performance and search counts within the given time intervals.

To uncover the relationship between the volume of search queries for a specific term and the overall direction of trader decisions, we analyzed adjusted closing prices and exchange volume changes on the stock market and used Google Trends to determine the number of searches that have been carried out for specific search terms, including industry-related words such as “fashion”, “footwear”, “beverage”, and companies’ signature products, such as “itunes”, “tazo”, or “coffee”. Next, we introduced and calculated the lag that’s necessary to create comparisons between the companies’ stock prices and their relevant search volumes. In this work, we provided a quantification of the relationship between changes in search volume and changes in stock market prices in hopes of finding trends in the stock market prices that correlate with the frequent usage of our companies’ keywords.

**2) Visualization**

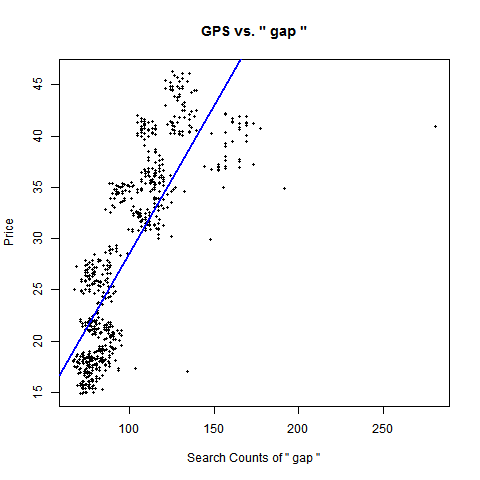
The variables we use in our analysis include stock price, exchange volume, number of searches (of our keywords), time, industry, lag (time for search counts to reflect stock price/volume change), and the correlations between search counts and stock price and volume. We made dot plots to observe changes in price, volume, and correlations over time, and compared the results across different industries by overlaying the dot plots or using bar charts.

**Results and Discussion**

**1) Individual Companies**

We discovered that there were some companies whose most frequently searched keyword was the companies’ names. Some of these companies include The Gap and Google, and their own stocks stood out as stellar examples of this finding. As seen below in Figures 1 and 2, the plots show that there seems to be a positive correlation between Google’s stock prices and the search counts for their company name “google”. This suggests that there is a positive correlation between a company’s stock market prices and the search counts of the company’s name in Google search, which may be due to increasing popularity with the companies’ products.

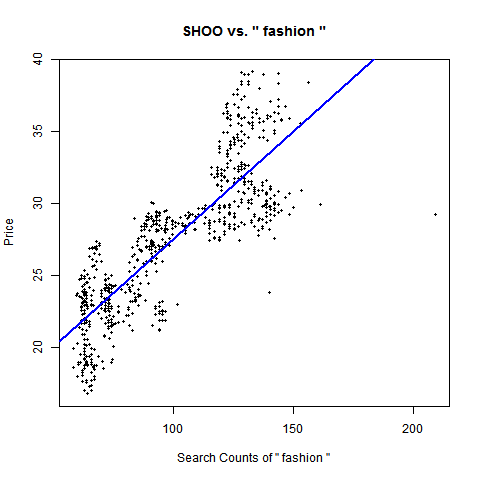
**Figure 1: GPS vs. Search counts of keyword “gap”**



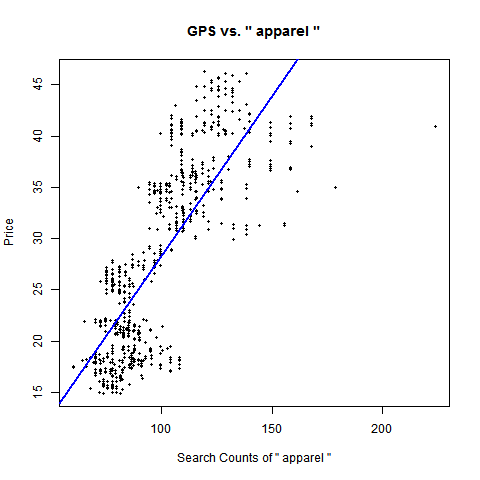
### Figure 2: GOOG vs. Search counts of keyword “google”

### 

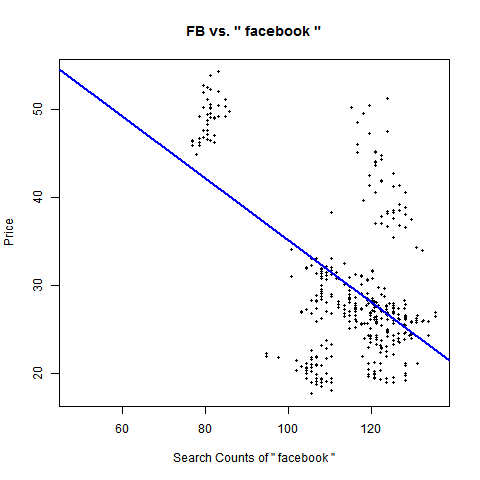
On the other hand, we found that some of the key words that show positive correlations are not always the companies’ names; industry-related terms are also well reflected on stock prices. For example, the keywords “fashion” for Steve Madden (Figures 3) and “apparel” for Gap (Figure 4) both show positive correlations between stock prices and search counts. The reason for this trend may be due to the fact that the industries that these companies belong to are becoming more and more popular and the companies themselves are growing. As a result, the search counts for industry-related keywords increases (due to rising interest) and more people buy the companies’ stocks, bringing the prices higher.

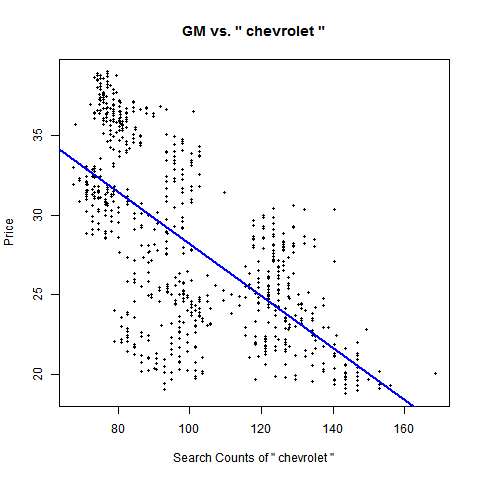
**Figure 3: SHOO stock prices vs. Search counts keyword “fashion”**

**Figure 4: GPS vs. Search counts of keyword “apparel”**

****

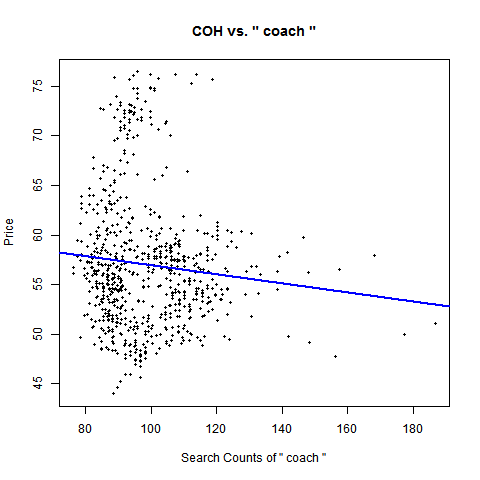
However, the correlations between the keywords and corresponding companies’ stock performances are not all positive. As seen below in Figures 5 and 6, there are negative correlations for the industry-related or company key words for GM and Facebook over the three years. Facebook’s stock price has dropped since its IPO in May of 2012. This could be interpreted as a case when drops in the financial market are preceded by periods of investor concern, hence investors may search for more information about the market before eventually deciding to buy or sell. On the other hand, there has been a downturn in the U.S. auto industry over the past few years. Since the Financial crisis of 2007-08, falling sales and market shares have resulted in the Big Three's plants’ underperformance, leading to production cuts, plant closures and layoffs. This trend continued until 2012 when GM and Ford continued to lose market shares as their sales went down.

**Figure 5: FB vs. Search counts of keyword “facebook”**

**Figure 6: GM vs. Search counts of “chevrolet”**

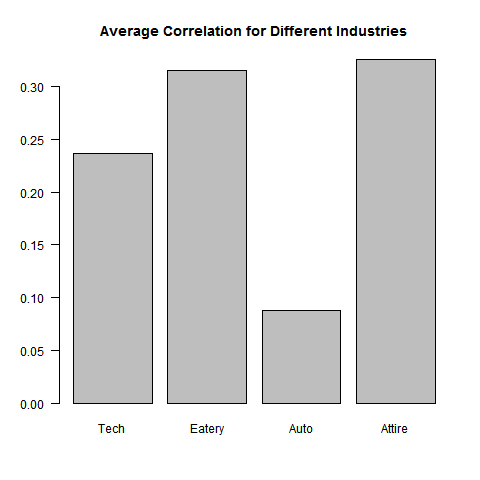
There are also cases where the correlation between the keyword and stock price is weak or nonexistent, despite the fact that the keyword may seem to be related to an important aspect of the company. An example can be seen below in Figure 7 where the keyword “Coach” actually correlates weakly with Coach’s stock prices.

**Figure 7: COH vs. Search counts of keyword “coach”.**



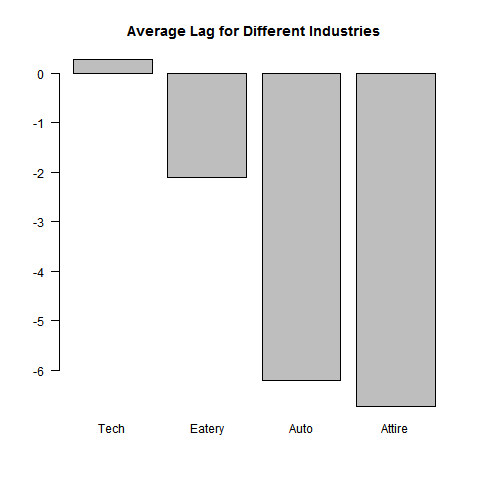
**2) Industry Comparison**

To better understand the effects of Googld search counts on a company’s stock performance, it makes sense to examine the differences across industries due to the differing nature of various businesses. We compiled the correlations between each company’s stock performances and search counts effects for the four different industries and created a bar chart below that shows the average correlation values among each industry.



As can be seen, clothing and eateries companies have the strongest correlation of approximately 0.30 between media exposure and stock performance. This is then followed by the tech industry with a correlation of roughly 0.24. Finally, the auto industry’s stocks stay indifferent of their media presence with a correlation coefficient < 0.1.

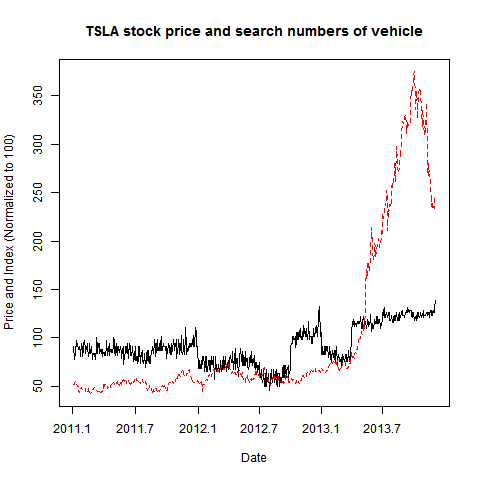
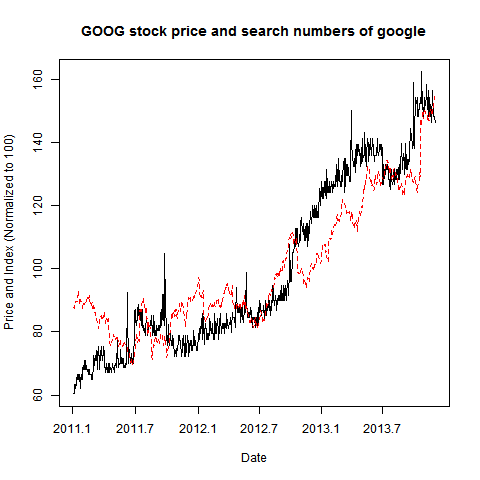
Taking the average lag between stock price and online searches into account, it seems that industries react differently to social media. The graph below illustrates the mean lag time (number of days) for different industries when averaging the responses in ±10 days. As we can see from the bar chart, tech firms’ stock prices move ahead of social media while eateries, automobile manufacturers and clothing companies’ stocks respond to social media with slight delays.



Firstly, the trend in lag indeed reinforces the fact that technology companies are coming back post the Dot-com bubble in 2000. After the Financial crisis of 2007-08, investors began pouring money into the market, leading to the increasing popularity of tech firms. However, the fact that tech firms’ stock prices move ahead of their social media exposure somehow suggests the irrationality of investors, the moves of stock price, which are driven by investors’ long/short positions, create media frenzy. Secondly, this also coincides with the nature of various industries. For instance, tech companies always show stronger presence in online media because they themselves are SNS, online advertisers, or sellers of related devices. In the meanwhile, people talk less about consumer related industries and more about tech industries.

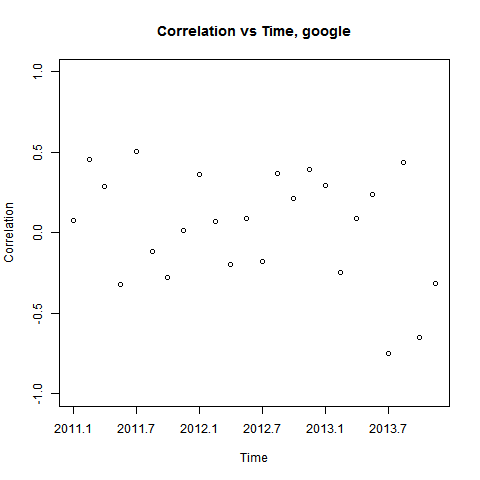
**3) Longitudinal Findings**

Since the Internet has been taking on a more integral role in people’s lives, it would be interesting to examine if there is any differences across time.



After normalizing both the stock price and search number to 100, most companies’ stock prices (indicated by red lines above) tend to move together with search numbers (indicated by black lines above) of a company’s most unique and relevant key word with a lag in time. Google (shown on the left graph above) demonstrates the most closely correlated trend over time. However, the graph of Tesla, shown on the right, stands out as the stock performance and search diverge after May 2013. This divergence is a possible indication that Tesla’s stock skyrocketed without any real gain for its investors.

It is still difficult to predict future stock performance solely based on the correlations. To better understand the time effect, we divide time into smaller units and analyze the correlation over time. If time does not play a role, then in each smaller unit where the effect of time is reduced, we should still observe consistent correlation between stock price and search counts.

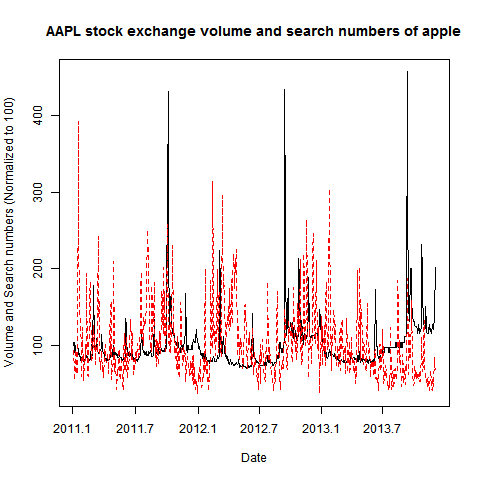


In the graph above, we used intervals of 30 days and plotted the correlations as data points and then drew the correlations for each single period in one graph. According to the graph, there is no specific pattern in the distribution of correlations, which shows that stock performance and media exposure lack a strong significant relation with one another. More plausibly, the positive correlation between stock price and search count is the result of them both increasing over time.

Overall, stock price tends to move together with social media exposure with a lag, and the correlation is not significant enough for people to take advantage of.

**4) Volume Impact**

Transaction volume may possibly have better correlation with media exposure compared to stock price because it has no direction.



In the graph above, transaction volume is indicated by the red line while the black line shows search numbers after normalizing to 100 for both axes. There is no identifiable pattern across time (due to many large spikes in volume and search numbers) and the trading volume is far more volatile compared to media exposure. The result shows that a stock’s daily transaction volume is not closely related to the company’s media exposure.

**Conclusions**

The search frequencies track what the stock prices could be. Our results suggest that, following this logic, Google Trends searches query volumes for certain terms that could have been used in the construction of profitable trading strategies. Where there are increases in search volume for certain company related terms, there were increases or decreases in stock prices for the respective companies as well. Whether the stock prices would increase or decrease can be determined with the context of the company conditions and economy itself—there are many other factors that could give more detailed insights about the companies’ futures. Also, there is variation in terms of response time and responsiveness across different industries. For example, technology companies tend to have active roles on online social media because of their business nature; meanwhile, thanks to the recent tech frenzy, their stock move volatilely without any public available information, which results in a low correlation between #search on Google Trends and stock performance. Over time, a company’s stock performance tend to move together with its online social media exposure, but after decomposing time into smaller periods, we found out that the correlation is in fact closer to a random distribution. The number of search might also affect a stock’s trading volume, but the stock market itself is quite unpredictable and volatile enough to weaken the correlation. Overall, it could be relatively difficult for us to predict, with 100% accuracy, the future stock prices solely from the search counts. Therefore, we could make the assumption that if there are some drops and peaks in search frequencies, then it is likely that there will be changes in stock prices.

Even though it may be lucrative to find out the relationships between a company’s stock performances with other factors, it could be quite challenging and unrealistic to predict where the market is heading to using any quantitative model. There may also be numerous confounding variables, including the economy’s conditions, overall stock market performance and investors’ sentiments, which all affect how a company’s stock performs.

**Reflection**

There are things we could improve upon. One big problem lies in key word selection. It is quite difficult to find 5 unique and representative key words for a company, especially when not all companies are like Apple, which owns many signature products. The word choices directly affect the Google Trends results that we obtained, thus influencing the correlation results we were analyzing.

**Anything to improve on?**

Even though it may be lucrative to find out the relationships between a company’s stock performances with other factors, it could be quite challenging and unrealistic to predict where the market is heading to using any quantitative model. There may also be numerous confounding variables, including the economy’s conditions, overall stock market performance and investors’ sentiments, which all affect how a company’s stock performs.