Can google trends data be used to design a profitable investment strategy?

Jugal Marfatia

EconS 529

Prof. J. Yoder

October 8, 2018

Significane of the paper:

Eugene Fama in his seminal paper on capital market efficiency proposed the efficient market hypothesis which states that all financial assets at any given time are marked correctly based on the existing available information [Malkiel and Fama, 1970]. The implication of Fama's paper is that there does not exist an arbitrage in the stock market and no systematic investment strategy can be designed to give higher returns than the overall stock market index in the long run. Since then many research papers have pointed out instances where markets are not correctly priced and there exists an arbitrage to exploit; however such arbitrage go away in the long run [Malkiel, 2003]. For example Malkiel (2003) mentions in his paper that researchers have pointed out significantly higher returns in the stock market during the first two weeks of January. However since the phenomena was documented the January effect has vanished. The main contribution of my paper will to design an investment strategy based on Google Trends data that will potentially have higher returns than the stock market index. This contribution will not only provide potential investors a systematic investment strategy to earn higher returns but it will also enhance the literature on the understanding of the efficient market hypothesis.

For my paper I propose to research two main things. First, study the correlation between Google Trends Index (GTI) for the stock ticker [1] and the price of the stock in the next period. In particular I will use the below two econometric models to study the correlation.

[1] GTI measures the relative search interest in a given period compared to other periods. GTI is a number between 1 and 100, where a number close to 1 indicates low search interest for the term in the given period and number close to 100 implies a high level of interest.

$$\%\Delta Price_{i,t+1} = \beta f(GT\_Index_{i,t}) + e_{it} \tag{1}$$

$$Prob(\%\Delta Price_{i,t+1} > 0) = \beta f(GT\_Index_{i,t}) + e_{it} \tag{2}$$

Where,

- $\%\Delta Price_{i,t+1}$, is the percentage change in the price of stock i in period $t+1$
- $GT\_Index_{i,t}$ is the GTI for stock i in period t
- $Prob(\%\Delta Price_{i,t+1} > 0)$ is the probability of percentage change in the price of stock i in period $t+1$ to be greater than 0.
- t is the time period and for the purpose of this paper is every week between 2013 and 2017.

Since there is no theory on the functional of $f(GT\_Index_{i,t})$ and the relationship between the two dependent variable and $GT\_Index_{i,t}$ is ambiguous, I will use the least absolute shrinkage and selection operator (LASSO) regression technique to determine the functional form of the relation and also estimate the corresponding coefficient parameters. A similar previous study estimated the correlation between the number of times a stock was mentioned in the financial times and the price of that stock in the next period [Alanyali et al., 2013]. Their results show that there is indeed a statistically significant relation, however the size and the sign itself of the relation is heterogeneous among different stocks. I expect similar results using the Google Trends data instead of financial times data.

The next contribution of my paper will be to use the results from the above regressions and design a profitable investment strategy. A previous study uses Google Trends data to design a investment strategy. The study argues that GTI can be used to minimize the standard deviation of the returns for a given portfolio [Kristoufek, 2013]. They weigh the stocks in their portfolio according to the below equation.

$$w_{i,t} = \frac{V_{i,t}^{-\alpha}}{\sum_{j=1}^{N} V_{j,t}^{-\alpha}} \tag{3}$$

Where, $w_{i,t}$ is the weight of stock i in time period t, $V_{i,t}$ is the GTI for stock i in period t and N is the number of stocks in the portfolio. In assigning weights to the stocks in their portfolio according to the above rule, the authors in the previous study do not assume any correlation between GTI and stock price, but rather find the optimal value for α that minimizes the variance of the portfolio return. For my paper I propose assigning weights to stocks in the portfolio according to the below rule (equation 4).

Number of Stocks	500 (S&P 500)
Time frame	01-01-2013 to 31-12-2017
Number of t (Weeks)	260
List of variables	Open/Close Daily Price, Daily Trading Volume, Weekly Google trends Index.

Table 1: Summary of the Data

$$w_{i,t} = \frac{pr_{i,t+1}}{\sum_{j=1}^{N} pr_{j,t+1}} \tag{4}$$

Where $pr_{i,t+1}$ is the probability that the price of stock i will increase in the next period. As suggested earlier, $pr_{i,t+1}$ will be estimated from equation (2).

Data

I have written a python script that automates the download process of the data. I have downloaded GTI and stock price data for all S&P 500 companies from 2013 to 2017. The code can be found in my github repository at this link. https://github.com/jugalm/Google-trends-stock-data/blob/master/fetch_data.ipynb.. The data can also found within the repository on this link https://github.com/jugalm/Google-trends-stock-data/tree/master/data.

References

[Alanyali et al., 2013] Alanyali, M., Moat, H. S., and Preis, T. (2013). Quantifying the relationship between financial news and the stock market. *Scientific reports*, 3:3578.

[Kristoufek, 2013] Kristoufek, L. (2013). Can google trends search queries contribute to risk diversification? *Scientific reports*, 3:2713.

[Malkiel, 2003] Malkiel, B. G. (2003). The efficient market hypothesis and its critics. *Journal of economic perspectives*, 17(1):59–82.

[Malkiel and Fama, 1970] Malkiel, B. G. and Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The journal of Finance*, 25(2):383–417.