Impact of Google Search Volume Index (GSVI) On Individual Stock Return

Yuqi Chen & Chan Hyuk Yoon

Introduction

Finding the factors behind the stock price movement have always been one of the biggest interests in financial research. Since the historical research of the Fama-French factor attribution model, the research horizon has widened to examine factors such as momentum, liquidity, and default risks. Following this trend, researchers are increasingly more focused on the impact of investor attention (Eirik, 2016). With a hypothetical approach that investor's increased attention on stock will have an impact on individual stock return, we utilize Google Trends to conduct linear regression. While discussion on whether search volume can forecast return is still ongoing (Preis, 2010), we hope to add value by (1) empirically testing recent data and (2) focusing on the individual stock rather than macro market or portfolio-wise price return.

Linear Regression Construction

The key questions and corresponding linear equation for empirical testing are as following:

1. Does the GSVI factor have explanatory power?

$$r_t = \text{Fama Three Factors} + \beta * GSVI_t + e_t$$

This is the basic regression model of our interest. r_t is a weekly stock return. Independent variables are Fama-French factors along with GSVI. GSVI is measured as the change of a particular keyword Google SVI divided by 100, where 100 is a scaling factor to make GSVI fall between 0 and 1. We run a regression with traditional three factors to examine if GSVI has any additional impact.

2. Does an increase in keyword searches have a lagging/momentum effect on returns?

$$r_t = \text{Fama Three Factors} + \beta_t * GSVI_t + \beta_{t-1} * GSVI_{t-1} + e_t$$

Based on prior research, an increase in keyword volume predicts higher stock prices in the next 2 weeks (ZHI DA & GAO 2011). We add in lagging variable to examine this effect.

3. Does increase in "positive" or "negative" keyword searches have an additional impact on returns?

$$r_t = \text{Fama Three Factors} + \beta_t * \text{adjusted } GSVI_t + \beta_{t-1} * \text{adjusted } GSVI_{t-1} + e_t$$

$$\text{where adjusted } GSVI = [\text{ positive word}(\Delta\%) - \text{negative word}(\Delta\%)\ ] * GSVI_t$$

An increase in keyword search volume could be due to either positive or negative news which will affect the firm's return differently. To take this directional aspect of GSVI into concern, we calculate and compare the percent changes of positive and negative words and multiply it with GSVI.

Data Construction

- ***Period Selection: Weekly Data from Jan 2014 ~ Present***
  - Among the selection between monthly, weekly, and daily data we choose a weekly period. This is because a month's period is too long to accurately assess the impact of event driven Google searches. On the other hand, there was a limitation in acquiring daily data for GSVI.

- ***Data Selection & Unit Explanation: Google Trends***
  - GSVI values are calculated in relatively. Values will appear as 100 for the most frequently searched search terms, 50 for search terms with half the frequency, and 0 if there is not enough data for that search term.
  - As mentioned, for our regression analysis we input percentage change of GSVI values as model input.

- ***Stock Selection: APPLE Company***

- We assume a company that qualifies following standards of (1) Popularity / Brand Recognition and (2) Trading volume are most likely to be searched online when retail investors are making an investment decision. Apple company is ranked the top with that standard.

- **Keyword Selection:** ***"AAPL", "AAPL Buy", "AAPL Sell"***
  - We use ticker rather than the company's name for Google keyword search. This is to differentiate between the literal meaning of fruit Apple. We also check how much keyword of ***"AAPL Buy"*** and ***"AAPL Sell"*** have been searched to discern investors' buying and selling intention when searching ticker ***"AAPL"***.

Descriptive Analysis

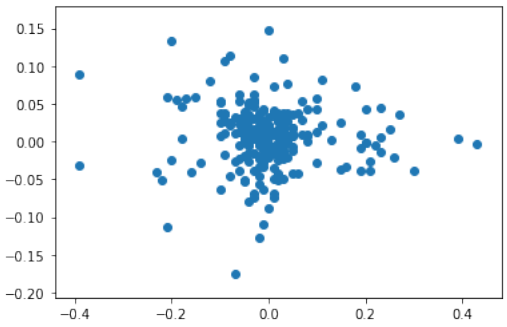
- **Descriptive features table**

	r	GSVIt	GSVIt_1	Mkt-RF	SMB	HML
count	244	244	244	244	244	244
mean	0.00635	0.00152	0.00164	0.00199	-0.00020	-0.00223
std	0.03913	0.09864	0.09733	0.02440	0.01251	0.01870
min	-0.17531	-0.39000	-0.39000	-0.14560	-0.05910	-0.08570
25%	-0.01289	-0.04000	-0.04000	-0.00615	-0.00783	-0.01080
50%	0.00825	-0.01000	-0.00500	0.00465	-0.00060	-0.00215
75%	0.02743	0.03000	0.03000	0.01463	0.00728	0.00660
max	0.14733	0.43000	0.43000	0.10310	0.06030	0.09870

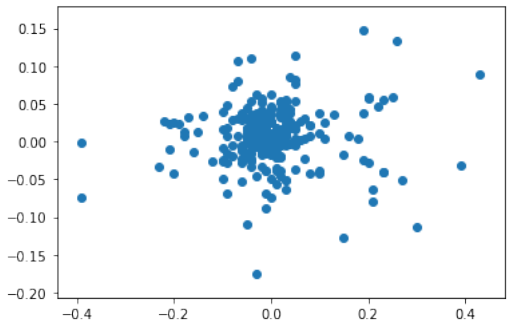
We run regressions based on the weekly data of the past 5 years since the week of October 30, 2015.

- **Scatter Plot**

Return and GSVI(t)



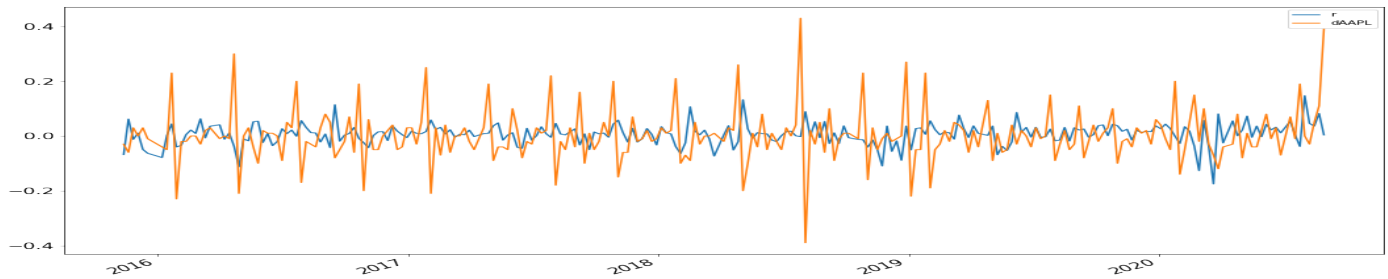
Return and GSVI(t-1)



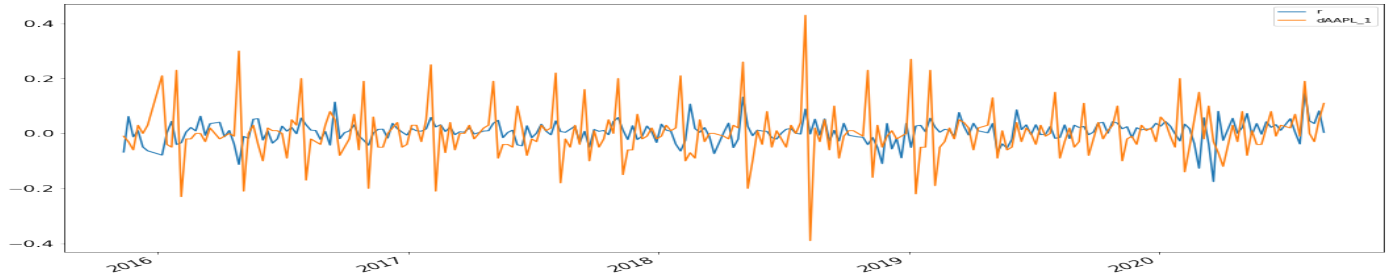
The pictures above are scatter plots of return of AAPL and GSVI at the same period and return of AAPL and GSVI at a week ago. We can see that there is not much correlation between the return of AAPL and GSVI of the same period, but there is a possible positive correlation of return of AAPL and lagged GSVI. This implies that investors' search act as a predictor of the stock's next week's return.

- **Time Series**

Return and $GSVI_t$



Return and $GSVI_{t-1}$



The pictures above are time series plot of return of AAPL and $GSVI_t$ and $GSVI_{t-1}$. In the first plot, we plot AAPL return and $GSVI_t$ together. We can observe that the orange line's ($GSVI_t$) peak is before the blue line's (AAPL return) peak, which means investors' attention increases first, and then stock return increases. In the second plot, we see a co-movement of AAPL return and $GSVI_{t-1}$.

Result Summary & Interpretation

Factors	R_squares	Adj_R_squares	AIC
['Mkt-RF']	0.415	0.412	-1012
['Mkt-RF', 'SMB', 'HML']	0.487	0.481	-1040
['Mkt-RF', 'SMB', 'HML', 'GSVI(t)']	0.497	0.489	-1043
['Mkt-RF', 'SMB', 'HML', 'GSVI(t-1)']	0.503	0.495	-1045
['Mkt-RF', 'SMB', 'HML', 'GSVI(t)','GSVI(t-1)']	0.507	0.496	-1045
['Mkt-RF', 'SMB', 'HML', 'GSVI*(t)', GSVI*(t-1)']	0.509	0.499	-1047

GSVI* = adjusted GSVI with concerning positive/negative connotation

1. Does GSVI factor have explanatory power? [Yes]

$$r_t = Fama\ Three\ Factors + \beta * GSVI_t + e_t$$

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.0029	0.002	1.573	0.117	-0.001	0.006
Mkt-RF	1.1612	0.077	15.013	0	1.009	1.314
SMB	-0.2721	0.15	-1.809	0.072	-0.568	0.024
HML	-0.5384	0.099	-5.433	0	-0.734	-0.343
GSVI(t)	-0.0419	0.019	-2.222	0.027	-0.079	-0.005

GSVI(t) factor can be said to have explanatory power for the following reasons :

- T-statistics of -2.22 and P > |t| = 0.027 means that we can reject null hypothesis of $\beta = 0$
- Both R squares and Adjusted R squares have increased while AIC decreased

2. Does an increase in keyword searches have a lagging/momentum effect on returns? [Yes and No]

$$r_t = Fama\ Three\ Factors + \beta_t * GSVI_t + \beta_{t-1} * GSVI_{t-1} + e_t$$

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.0027	0.002	1.471	0.143	-0.001	0.006
Mkt-RF	1.1708	0.077	15.22	0	1.019	1.322
SMB	-0.2458	0.15	-1.641	0.102	-0.541	0.049
HML	-0.5626	0.099	-5.68	0	-0.758	-0.367
GSVI(t)	-0.0271	0.02	-1.357	0.176	-0.066	0.012

GSVI(t-1)	0.0436	0.021	2.119	0.035	0.003	0.084

Adding **GSVI(t-1)** factor can be said to have explanatory power for the following reasons:

- T-statistics of -2.119 and P > |t| = 0.035 means that we can reject null hypothesis of $\beta = 0$
- Both R squares and Adjusted R squares have increased while AIC decreased

Note that the coefficient is positive for **GSVI(t-1)**. The reasons could be as following:

- Since there is a time between searching the information and making the decision, increased search a week ago could be reflected in the current week's price.
- This implies the possibility of a **lagging effect of the GSVI variable.**

Note that coefficients for **GSVI(t) and GSVI(t-1)** are the opposite.

- If there was a momentum effect between time t and t-1, both would have the same sign of coefficients. Though the relationship would be more clear as we add in more variable of t-2, time-series graph of GSVI and return displays weekly cyclic movement.
- This implies there might be a possibility of **no momentum effect of the GSVI variable.**

3. Does considering "positive" or "negative" keyword searches have an additional impact? [Soso]

$$r_t = \text{Fama Three Factors} + \beta_t * \text{adjusted } GSVI_t + \beta_{t-1} * \text{adjusted } GSVI_{t-1} + e_t$$

$$\text{where adjusted } GSVI = [\text{ positive word}(\Delta\%) - \text{negative word}(\Delta\%)\ ] * GSVI_t$$

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
const	0.0027	0.002	1.516	0.131	-0.001	0.006
Mkt-RF	1.1602	0.077	15.134	0	1.009	1.311
SMB	-0.2385	0.149	-1.596	0.112	-0.533	0.056
HML	-0.532	0.099	-5.375	0	-0.727	-0.337
Adj_GSVI(t)	0.0706	0.024	2.912	0.004	0.023	0.118
Adj_GSVI(t-1)	0.0673	0.028	2.405	0.017	0.012	0.122

Changing to **Adjusted GSVI** factors do not have a critical impact on regression improvements:

- Both R squares and Adjusted R squares have increased while AIC decreased but only slightly
- However, confidence to reject null hypothesis of $\beta = 0$ is increased when we consider directional movement of increment in stock search volume. T-statistics of -2.912 and 2.405, P > |t| = 0.004 and 0.017 means that we can reject null hypothesis of $\beta = 0$ with confidence.

The reason behind only a slight improvement in regression could be the following:

- This could be because retail investors usually search on google with the "buying" intention rather than "selling" intention. This is because of asymmetry in the decision-making environment. When an investor buys a stock, he or she has to select a stock among diverse stock. On the other hand, when an investor sells a stock, they have no other choice to sell from what they own. This asymmetry might incline an investor to search when buying the stock. In turn, this would dissolve the effect of *[positive word(Δ%) – negative word(Δ%)]* adjustment, for negative word changes would be neglectable.

References

- Laurens Bijl, Glenn Kringhaug, Peter Molnár, Eirik Sandvik (2016) Google searches and stock returns, International Review of Financial Analysis 45 150–156
- Preis, T., D., & Stanley, H. E. (2010). Complex dynamics of our economic life on different scales: insights from search engine query data.
- Zhi Da, Joseph Engelberg, and Pengjie Gao (2011). In Search of Attention. THE JOURNAL OF FINANCE • VOL. LXVI, NO. 5