Google Searches and Stock Returns

Neha Aggarwal

April,2020

1 Key Objective

To study the effect of Google Search Volumes (GSV) on stock returns.

2 Research Question

Can search query data on company names be used to predict weekly stock returns for individual firms ?

3 Data Used

They used two data sources, one is Wharton Research Data Services (WRDS) and the other one is Google Trends. The data (from January1, 2007 through December 31, 2013) regarding daily open prices, volumes, dividends and the number of shares for companies in the S&P 500 index was obtained from WRDS. The GSV data was taken from Google Trends. The GSV data was used in the form of indices (with values from 0 to 100) for the search volumes in US for the names of companies in the S&P 500. Only those companies whose complete stock data back to 2007 was available (i.e. those companies which were part of S&P 500 index for the whole period of 2007 - 2013), were considered for their data set. This left them with a complete dataset on 431 companies.

4 Sample Period

The GSV data was analysed from 2008 to 2013 but the stock data was analysed from 2007, in order to calculate betas and averages for stocks in 2008.

5 Methodology

The dataset used was a panel with 6 years (313 weeks) of observation for 431 companies.Therefore, panel data regressions was used. So, firstly lets try to explain our regression variables.

5.1 Regression Variables

5.1.1 Excess Stock Return

Excess return is calculated by subtracting the beta of the individual stock multiplied by market returns from daily stock returns. Daily stock return is calculated as total return adjusted for dividends and stock splits as shown below, where S is stock open price,D is dividend,N is the number of shares outstanding,t is time in days, $R_{d,t} = \frac{(S_t+D_t)N_t}{S_{t-1}N_{t-1}} - 1$
Weekly stock return is given by - $R_{w,t} = \prod_{i=1}^{n}(1 + R_{d,i}) - 1$
Weekly excess return is given as - $R_t = R_{w,t} - \beta R_{M,W,t}$

5.1.2 Standardized GSV

The indices from Google Trends were used to calculate a standardized GSV. Standardization was done to make these indices more comparable across companies. The equation of standardized GSV is as given below, where n is the number of weeks of GSV observations, and σ_{GSV} is the full-sample standard deviation of the GSV time series. $SGSV_t = \frac{GSV_t - \frac{1}{n}\sum_{i=1}^{n} GSV_i}{\sigma_{GSV}}$

5.1.3 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. $\sigma_{w,t} = (\sum_{i=1}^{n} r_d^2)^{\frac{1}{2}}$
This is medium term volatility (weekly). For long term volatility (monthly), we simply average the weekly volatilities for the last 5 weeks. $\sigma_{l,t} = \frac{1}{5}\sum_{i=t-4}^{t}\sigma_{w,i}$

5.1.4 Trading Volume

In this model detrended log volume (denoted as Vlm) is used, where the trend is a rolling average of the past 12 weeks of log volume. $Vlm_t = log(Volume_t) - \frac{1}{12}\sum_{i=t-11}^{t} log(Volume_i)$

5.2 Regression Models

The explanatory variables used are five lags each of excess volume (Vlm), excess return (R) and the standardized GSV (SGSV). The lag operator is denoted as L.
So, our first model is- $R_t = \alpha + (\sum_{i=1}^{5}\beta_i L^i)R_t + (\sum_{i=1}^{5}\gamma_i L^i)SGSV_t + (\sum_{i=1}^{5}\delta_i L^i)Vlm_t + \theta\sigma_{w,t-1} + \xi\sigma_{l,t-1} + \epsilon_t$

In our next model, we also include January effect variable (Jan) and three interaction variables ($Vlm \times R, R \times SGSV, Vlm \times SGSV$). So the extended model becomes-
$R_t = \alpha + (\sum_{i=1}^{5}\beta_i L^i)R_t + (\sum_{i=1}^{5}\gamma_i L^i)SGSV_t + (\sum_{i=1}^{5}\delta_i L^i)Vlm_t + \theta\sigma_{w,t-1} + \xi\sigma_{l,t-1} + \epsilon_t + \tau(Jan) + \lambda(Crisis) + \rho(R_{t-1}SGSV_{t-1}) + (Vlm_{t-1}SGSV_{t-1}) + \omega(R_{t-1}Vlm_{t-1})$
where all greek letters are regression coefficients.

6 Major Results

Results are presented as the impact on excess return (in basis points) of a one standard deviation change in the independent variable. Absolute values of coefficients are not used because our all variables are not standardised and thus the size of the coefficients depends on the scale of the variables.
After examining the effect of SGSV, the following results come up-

- The effect of the standardized GSV from the previous week (1 lag) is consistently negative and significant across both models. This effect seems stronger after the financial crisis(2010-2013). For the two-week lag positive effect is weaker. This effect is weaker for the post-crisis period. The greatest impact in basis points occurs three weeks after a change in GSV. However,that effect is relatively small and is not significant in the post-crisis period.
- The predictive power of GSV is higher than the effect of the detrended volume. However, compared to the lagged excess returns and volatility the impact of GSV is weaker.
- The inclusion of interaction and crisis/January dummy variables in our model does not change the impact of GSV on excess return.
- Comparing the regression results with and without the financial crisis, we observe that the auto-correlation in the excess returns has been significantly reduced. This is expected, as market crashes often lead to increased auto-correlation in stock returns.
- In addition to the excess return used throughout $R_{w-\beta m}$, ordinary weekly return R_w and excess returns calculated as regular weekly return minus market return R_{w-m}, are also used to test the robustness of the model. These different definitions of returns generally yield similar and significant results even though they are weaker. This supports our hypothesis that GSV predicts stock returns.

7 Conclusion

- A regression was tested on absolute excess returns of GSV alone and significant, positive coefficients were found. Although a simple regression would indicate a positive correlation, other variables(especially volatility) are better at explaining the magnitude of returns than GSV.
- High levels of GSV predict low future excess returns. The coefficients of the GSV variables are statistically significant,but their impact is small. The predictive power of GSV is similar both during the financial crisis and in more ordinary market conditions.
- A trading strategy based on selling stocks with high Google search volumes and buying stocks with infrequent Google searches was also examined. This strategy is profitable when the transaction cost is not taken into account but is not profitable if we take into account transaction costs.
- Results confirm that Google search volumes can predict stock returns. However,the relationship between Google search volumes and stock return changes over time.