ONLINE SEARCH: IDENTIFYING NEW INVESTMENT HABITATS

Completed Research Paper

Alvin Leung
The University of Texas at Austin
2110 Speedway Stop B6500
Austin TX 78712
aleung@utexas.edu

Ashish Agarwal
The University of Texas at Austin
2110 Speedway Stop B6500
Austin TX 78712
Ashish.Agarwal@mccombs.utexas.edu

Prabhudev Konana
The University of Texas at Austin
2110 Speedway Stop B6500
Austin TX 78712
Prabhudev.Konana@mccombs.utexas.edu

Alok Kumar
The University of Miami
School of Business Administration,
514 Jenkins Building,
Coral Gables, FL 33124
akumar@bus.miami.edu

Abstract

Researchers often refer to investment habitats/categories to explain the patterns of co-movement in asset returns that cannot be fully clarified by fundamentals. Many factors determine these habitats including investor preferences to size, industry, price-levels and risk-levels. This paper investigates a unique method to explore investment habitat based on the search behavior of investors on the Internet without actually using proprietary trading data. Using Yahoo! Finance data on investors' frequently co-viewing stocks of Russell 3000 stocks between September 15, 2011 and February 24, 2012, we evaluate the return co-movement within investment search habitats/clusters. We find that stocks within a search cluster show strong co-movement with other stocks in the same cluster that is not fully explained by other traditional habitat characteristics. The behavior persists even when a stock moves from one cluster to another. The study provides direct empirical evidence that investors prefer to search among stocks that have similar co-movement.

Keywords: Comovement, graph theory, investment habitat, market, value, search behaviors

Online Search: Identifying New Investment Habitats

Completed Research Paper

Introduction

Analyzing ex post proprietary trading data, researchers have uncovered comovement patterns in returns of a subset of stocks. Researchers attribute comovement to traditional stock fundamentals related to cash flows, and a host of other factors based on investor preferred habitats. In this habitat-based model, investors with certain attributes buy or sell a subset of stocks in concert with others resulting in systematic comovements (Barberis et al. 2005; Kumar and Lee 2006). Some factors leading to investment habitats include traditional measures such as company size and institutional ownership (Froot and Teo 2008; Pindyck and Rotemberg 1993); inclusion in an index such as S&P 500 (Barberis et al. 2005; Vijh 1994) and Nikkei 225 (Greenwood 2008); lottery features such as low-prices, high volatility, and high skewness (Kumar 2009); and stock splits (Green and Hwang 2009; Kumar et al. 2012). This research explores whether extracting investors' online search behavior can uncover investment habitats without using proprietary trading data.

Millions of individual investors search for stock information and participate on message boards over the Internet leaving a digital trail. This trail has made it possible to investigate online investor behaviors and their impact on the stock market performance. For example, researchers find that the message volume and sentiments expressed on message boards can predict market returns, trading volume, and market volatility (Antweiler and Frank 2004; Das and Chen 2007). Da et al., (2011) show that online user search behavior for individual stocks is a good predictor of the market returns of those stocks in the next two weeks. This stream of literature has focused on predicting individual stock returns and has not explored potential investment search habitats with return comovements.

Da et al. (2011) reveal that online search for a stock indicates an active interest and the investors are very likely to invest in that stock. We can conjecture then if investors share common interest in a group of stocks, then their search behavior should also reveal such groups or clusters (henceforth, stocks traded or viewed together is called a cluster). Prior literature supports this argument that investors tend to focus on a group of stocks that have common attributes because of limited investor attention resulting from scarce cognitive resources. Peng et al. (2007) argue that because of limited investor attention, investors tend to allocate more attention to market or sector-level common factors– referred to as category learning behavior – rather than to factors that are firm specific. If this is the case then search behavior of stocks should reveal investors' preferences to a group of stocks – that is, search clusters – with common factors including, but not limited to, market or sector-level factors. It is likely that search clusters may represent investors' investment habitats and stocks in the cluster show return comovement. Thus, in this research we address two questions: (a) Does investor search behavior reveal search clusters or investment habitat?; (b) Do the stock returns in the same search cluster move together?

Identifying search clusters has several advantages over traditional ex-post analysis using proprietary trading data. First, popular finance portals like Yahoo! Finance provides stocks that are viewed together. Hence, clusters can be extracted on a continuous basis using broader market participants. Second, search clusters may incorporate pre-trade search activity, post-trade monitoring activity, and curiosity. If search reflect pre-trade activity or curiosity we can understand the evolving investment habitats and changing habitats. Further, rather than finding clusters that are based on single factor like growth rates, size, and industry sector, the search clusters represent aggregated belief of investors based on various dimensions.

We use Yahoo! Finance search data to analyze user search behaviors for the Russell 3000 stocks. Yahoo! Finance provides a list of top also-viewed stocks for every stock (Figure 1). We use this co-viewing data and apply graph theory to identify commonly viewed search clusters. We then study the returns of each stock in each cluster as a function of the average return of the cluster excluding the focal stock. We control for fundamental risk factors such as market return, Fama-French 49 industry factors, and Carhart 4 factors. We also control for the effect of stock mention in the popular news in the same search cluster. We find positive comovement among stocks in the same search cluster. This suggests that online user search data can be used to identity investor habitats.

We investigated comovement of stock returns when a stock's cluster membership changes. We find that when a stock is added (removed) to a cluster that stock returns shows higher (lower) comovement with the cluster. This behavior persists when a stock switches from one cluster to another; that is, a stock shows a higher comovement with the cluster it belongs to at a particular point in time as compared to the comovement with the other cluster.

We also analyze the known characteristics (e.g. size, value, and industry) of stocks in a cluster that can potentially explain comovement. Apart from the traditional characteristics, we find that stocks also share high similarity in price volatility and supply-chain relationship. We then investigate the comovement of each stock with an alternative matched cluster of stocks based on Fama-French 10 industry category and similar market capitalization and price-to-book ratio (Massa and Zhang 2009). We find that the stocks have inferior comovement with the alternative matched cluster as compared to the search cluster. This implies that stocks in a search cluster possess other common characteristics apart from being similar in market capitalization (size), price-to-book ratio (value), and industry. Risk preference (volatility) and supply-chain relationship may cause investors to trade some stocks more frequently and this leads to the positive comovment in some of the search clusters.

This research contributes to the existing stream of literature on comovement and market sentiments. Prior work has focused on establishing the concept of investment habitats and demonstrated comovement of stocks in these habitats. We add to this literature by showing that online search data can be used to identify investment habitats. We also show that these investment habitats cannot be fully explained by the traditional characteristics of size, industry type, and value used to classify stocks. Our research also contributes to the existing literature on online search and its impact on market performance. Prior research has focused on user search behaviors and outcomes in the auto market (Kuruzovich et al. 2008), housing demand (Wu and Brynjolfsson 2009), and market returns of individual stocks (Da et al. 2011). This research extends to group of stocks. We show that that investor search data can be a useful source to determine the investment habitats.

Data and Research Model

We focus on the search behavior for Russell 3000 stocks used in previous literature ((Da et al. 2011; Diether et al. 2009; Evans et al. 2009; Haugen and Baker 1996). Russell 3000 comprises of firms whose market capitalization accounts for 98% of total market capitalization of all stocks trading in US exchanges. We collected user search data for the above stocks from Yahoo! Finance website. Yahoo! Finance is one of the most popular investment portals among individual retailers and it ranks consistently number one in terms of the popularity and the number of visits.[1]

On Yahoo! Finance, when users search for a particular stock say Bank of America (BAC) it also shows six other stocks that users have commonly viewed along with the BAC as shown in Figure 1. Yahoo! computes the co-viewed data based on visitors' cookies[2] and upload the most recent data to Yahoo! Finance.[3]

We wrote a program to collect daily co-viewing data for all stocks in the Russell 3000 index from September 15, 2011 to February, 24 2012. Our final sample consists of data for 2,900 individual stock data as some stocks were delisted during the period of research. We obtained the return data for each stock during this period from the CRSP database. We also collected the daily news data for our sample stocks from Google Finance.

---

[1] Top 15 popular business websites: http://www.ebizmba.com/articles/business-websites
Top 10 financial news and research websites:
http://www.comscore.com/Press_Events/Press_Releases/2008/07/Yahoo!_Finance_Top_Financial_News_and_Research_Site_in_US

[2] Cookies allow a website to identify and track all user activities including search for different items (in our case stocks).

[3] We have separately verified the data generation process directly with customer service of Yahoo! Finance

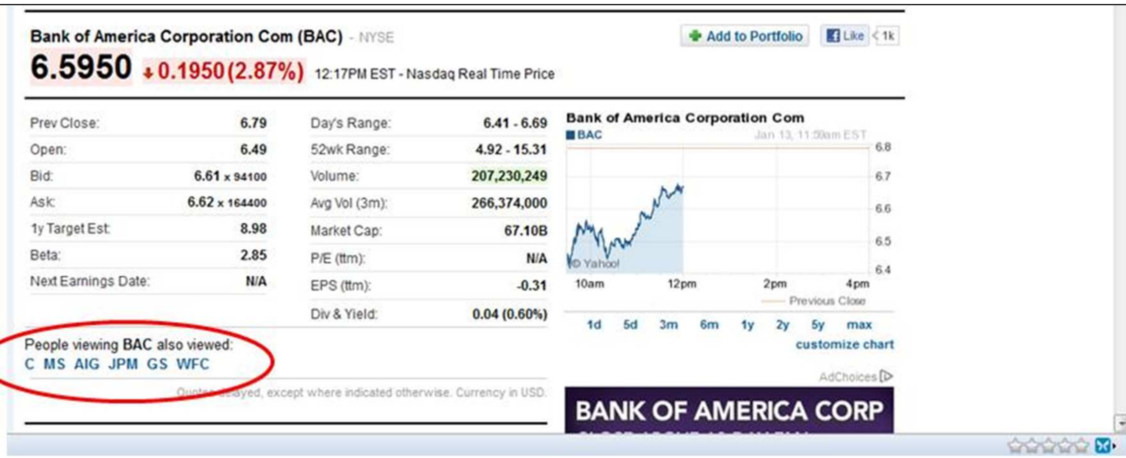

Figure 1. Example of Yahoo! Finance co-viewing data

Search Clusters

We determine search clusters based on the first six weeks (September 15 to October 31, 2011) of Yahoo! Finance co-viewing data. We form a directed search graph based on the stock data collected. Each vertex represents a stock ticker and a directed edge connects from each stock that has co-viewing data to individual stocks that have been viewed. The weight of each directed edge is the number of days such co-viewing pattern exists. To ensure that the sample data are persistent over time, we filter out directed edges if their weight is less than three weeks.

Later, we use the filtered data to determine cliques, which are the most fundamental unit of search clusters. According to graph theory, a clique is a sub-graph with at least three members in which all members are linked to each other. In our research context, we define a search clique to be a sub-graph where each vertex has at least three bi-directional edges with all other vertices. In our sample data, AIG, BAC, and C form a clique because each company appears on the others' co-viewing list. As some cliques may have overlapping members with the peers. We join cliques with common members together to form a bigger cluster. In this way, all clusters do not have any overlapping elements. Based on the data we have collected in the initial month, we identify 73 search clusters that comprise of a total of 318 stocks. Due to unavailability of some stock data, 8 stocks are removed from the analysis. Only 72 clusters and 310 stocks are retained.[4] The 310 stocks have a total market capitalization of $6.56 trillion on October 31, 2011 representing approximately one third of total market capitalization of all stocks in US stock exchanges.

Analysis of Comovement

We use stock returns after November 1, 2011 for the following analysis. The research models used in the analysis are similar to Barberis et al. (2005), who measure the comovement of stocks before and after inclusion in S&P 500.

To test whether there exists search cluster comovement, we run (1) with control of market returns:

$$R_{it} = \beta_{0i} + \beta_{1i} R_{ct} + \beta_{2i} R_{ct-1} + \beta_{3i} R_{S\&Pt} + \beta_{4i} R_{S\&Pt-1} + \beta_{5i} News7_{it} + \beta_{6i} News14_{it} + \varepsilon_{it} \tag{1}$$

Also, we add controls such as Fama-French 49 industry factor, Carhart four-factor, and aggregated news in the same cluster in the main model as below:

[4] Some stocks do not report their shareholder's equity and thus we cannot compute their price-to-book ratio. Also, some stocks are listed in the analysis period but delisted in the subsequent period of comovement analysis. As a result, those stocks are removed from the analysis.

$$R_{it} = \beta_{0i} + \beta_{1i}R_{ct} + \beta_{2i}R_{ct-1} + \beta_{3i}FF\_49_{it} + \beta_{4i}(R_{mt} - R_{ft}) + \beta_{5i}SMB_t + \beta_{6i}HML_t + \beta_{7i}UMD_t + \beta_{8i}News7_{it} + \beta_{9i}News14_{it} + \varepsilon_{it} \quad (2)$$

where

R_{it} is the return of stock i at time t

R_{ct} is the average capitalization weighted return of cluster c in which stock i belongs to at time t excluding the return of stock i

R_{ct-1} is the lagged average capitalization weighted return of cluster c

$R_{S\&Pt}$ is the average capitalization weighted return of S&P 500 at time t

$R_{S\&Pt-1}$ is the lagged capitalization weighted average return of S&P 500

FF_49_{jt} is the Fama-French 49 industry return firm j belong to at time t

R_{mt} is CRSP value weighted market portfolio return at time t

R_{ft} is the risk-free rate at time t

SMB_t is the Fama-French size factor at time t

HML_t is the Fama-French book-to-market factor at time t

UMD_t is the Carhart momentum factor at time t

$News7_{it}$ is the log value of total number of news of the cluster in which stock i belongs to in the most recent 7 days (i.e. t, t-1, ..., t-6) plus 1

$News14_{it}$ is the log value of total number of news of cluster in which stock i belongs to in the next most recent 7 days (i.e. t-7, t-8, ..., t-13) plus 1

t is the time period when Yahoo search data are collected

FF_49_{jt}, R_{mt}, R_{ft}, SMB_t, HML_t, and UMD_t are retrieved from Fama-French data library (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html).

To control for autocorrelation, we run the research models as shown above using Newey-West standard error. To account for cross-sectional dependence, we use two approaches to compute the level of significance. First, we use asymptotic theory to determine robust standard errors. Second, as a robustness check, we follow the approach of Boyer (2011) and compute the p-value using block bootstrap samples. We determine the optimal block size based on our sample data using the method suggested by Politis and White (2004) and Patton et al. (2009). Then we draw a block of cross-sectional data with optimal block size randomly from the sample data with replacement. The bootstrap block is stacked in a sequential order and is used in the regression models as specified above. The procedure is repeated 10,000 times. If the estimated coefficient is positive (negative), the bootstrap p-value is computed as the proportion of average beta greater (smaller) than zero in the 10,000 bootstrap samples.

Results

Analysis of comovement

Table 1 show the summary statistics of all variables in different time period. With control of market returns and other macroeconomic factors, Tables 2 and 3 show that stocks in general demonstrate a significant and positive search cluster comovement. The magnitude of average search cluster betas, $\bar{R}_{C_it}$, is 0.61 and 0.31 in models (1) and (2), respectively. Both betas are significant at 1%. The results suggest that stocks within the same search cluster tend to comove together. The results are consistent with the category/habitat view of stock comovement (Barberis et al. 2005). Because users view some stocks together consistently, they form a habitat. Market information that changes the habitat may diffuse faster among stocks in the same habitat and such information is incorporated into the price of individual stocks more quickly.

Table 1. Summary Statistics														
	R_{it}	R_{ct}	R_{ct-1}	R_{pt}	R_{pt-1}	$R_{S\&Pt}$	$R_{S\&Pt-1}$	$News7_{it}$	$News14_{it}$	$FF49_{it}$	R_{mt}-R_{ft}	SMB_t	HML_t	UMD_t
Panel A. Full sample data (Nov 1, 2011 to Feb 24, 2012)														
N (firm-days)	24,490	24,490	24,490	24,490	24,490	24,490	24,490	24,490	24,490					
Mean	0.0016	0.0015	0.0011	0.0015	0.0011	0.0012	0.0008	0.4176	0.3636					
SD	0.03	0.02	0.02	0.02	0.02	0.01	0.01	0.47	0.42					
Min	-0.41	-0.26	-0.26	-0.08	-0.08	-0.04	-0.04	0.00	0.00					
Max	0.78	0.40	0.40	0.11	0.11	0.04	0.04	1.92	1.77					
Panel B. Subsample data (Nov 1, 2011 to Dec 31, 2011)														
N (firm-days)	13,020	13,020	13,020	13,020	13,020			13,020	13,020	13,020	13,020	13,020	13,020	13,020
Mean	-0.0006	-0.0006	-0.0013	0.0002	-0.0004			0.2979	0.3309	0.0001	0.0001	-0.0001	0.0004	0.0016
SD	0.03	0.03	0.03	0.02	0.02			0.34	0.36	0.02	0.02	0.01	0.0049	0.01
Min	-0.41	-0.26	-0.26	-0.08	-0.08			0.00	0.00	-0.07	-0.04	-0.01	-0.01	-0.02
Max	0.24	0.16	0.16	0.11	0.11			1.32	1.34	0.12	0.04	0.02	0.01	0.02

Table 2. Comovement Analysis with Search Cluster Return and Market Return						
Panel A. Summary Statistics of Betas						
	$R_{C_i t}$	$R_{C_i t-1}$	$R_{S\&Pt}$	$R_{S\&Pt-1}$	$News7_{C_i t}$	$News14_{C_i t}$
Mean	0.61	-0.001	0.56	0.03	-0.0004	-0.00004
SD	0.37	0.22	0.54	0.40	0.0098	0.01060
Min	-0.54	-1.51	-1.06	-2.00	-0.05	-0.06
Max	2.00	0.88	2.42	2.89	0.08	0.06
Panel B. Regression Results						
	$\bar{R}_{Ct}$	$\bar{R}_{Ct-1}$	$\bar{R}_{S\&Pt}$	$\bar{R}_{S\&Pt-1}$	$\overline{News7}_{Ct}$	$\overline{News14}_{Ct-1}$
Coefficient (Robust SE)	0.61 (0.01)	-0.001 (0.01)	0.56 (0.028)	0.03 (0.03)	-0.0004 (0.0006)	-0.00004 (0.0007)
Bootstrap p-value	0.00	0.48	0.00	0.13	0.12	0.57

Furthermore, Table 2 shows that the average cluster beta $\bar{R}_{C_it}$(0.61) is higher in magnitude than market beta $\bar{R}_{S\&Pt}$(0.56). In Table 3, with presence of Fama-French industry return and Carhart 4 factors, the significant and positive comovement of search cluster still persists. The average cluster beta $\bar{R}_{C_it}$ is significant whereas the average market beta $\overline{R_{mt} - R_{ft}}$ is insignificant with bootstrap p-value above 0.05.

In addition, we find that the lag average betas of $\overline{News7}_{Ct}$ and $\overline{News14}_{Ct}$ in Tables 3 and 4 are all insignificant. The results suggest that news have limited explanatory power in explaining individual stock movement.

As a robustness check, we also use equally weighted cluster returns instead of capitalization weighted cluster returns in (1) and (2) and replace weekly news by daily news. The results are qualitatively similar.

Table 3. Comovement Analysis with Search Cluster Return, Fama-French and Carhart Factors

Panel A. Summary Statistics of Betas

	R_{C_it}	R_{C_it-1}	$FF49_{it}$	$R_{mt} - R_{ft}$	SMB_t	HML_t	UMD_t	$News7_{C_it}$	$News14_{C_it}$
Mean	0.31	0.01	0.80	-0.14	0.13	-0.14	-0.35	-0.0024	-0.0009
SD	0.50	0.14	1.28	1.27	0.93	1.14	1.24	0.02	0.02
Min	-1.56	-0.88	-4.19	-10.90	-2.78	-6.43	-8.81	-0.15	-0.28
Max	1.78	0.49	10.30	6.48	4.65	5.22	6.03	0.06	0.07

Panel B: Regression Results

	$\bar{R}_{Ct}$	$\bar{R}_{Ct-1}$	$\overline{FF49}_t$	$\overline{R_{mt} - R_{ft}}$	$\overline{SMB}_t$	$\overline{HML}_t$	$\overline{UMD}_t$	$\overline{News7}_{Ct}$	$\overline{News14}_{Ct}$
Coefficient (Robust SE)	0.31 (0.02)	0.01 (0.01)	0.80 (0.05)	-0.14 (0.06)	0.13 (0.06)	-0.14 (0.06)	-0.35 (0.06)	-0.0024 (0.0010)	-0.0009 (0.0010)
Bootstrap p-value	0.00	0.18	0.00	0.09	0.01	0.01	0.00	0.19	0.22

Placebo test

To investigate whether the positive and significant within-cluster comovement is due to factors such as size, value, and industry, we conduct a placebo test by matching each member of a cluster by a corresponding stock within the same industry and with similar size and value. We follow similar matching methodology as Massa and Zhang (2009) who analyze the effect of style investing on mergers. First, based on all stocks in CRSP with valid market capitalization, price-to-book ratio, and SIC, we find stocks that fall into the same Fama-French 10 industry category for each individual member of a cluster. The stocks in the same cluster or being matched to another stock in the same cluster are not considered. Second, we find the absolute differences in market capitalization and price-to-book ratio between the firms identified in the previous step and the individual cluster member. Third, we rank those firms independently according to the absolute differences. Fourth, we sum up the ranks of the two absolute differences and choose the firm with the smallest sum of ranks. If there are two or more firms with equal sum of ranks, we choose the one with smallest absolute difference in market capitalization as matched stocks. The resultant matched stocks form a placebo cluster of each search cluster.

Upon identifying the corresponding placebo clusters, we run the following regression models:

$$R_{it} = \beta_0 + \beta_1 R_{ct} + \beta_2 R_{ct-1} + \beta_3 R_{pt} + \beta_4 R_{pt-1} + \beta_5 R_{S\&Pt} + \beta_6 R_{S\&Pt-1} + \beta_7 News7_{it} + \beta_8 News14_{it} + \varepsilon_{it} \quad (3)$$

$$R_{it} = \beta_0 + \beta_1 R_{ct} + \beta_2 R_{ct-1} + \beta_3 R_{pt} + \beta_4 R_{pt-1} + \beta_5 FF49_{it} + \beta_6 (R_{mt} - R_{ft}) + \beta_7 SMB_t + \beta_8 HML_t + \beta_9 UMD_t + \beta_{10} News7_{it} + \beta_{11} News14_{it} + \varepsilon_{it} \quad (4)$$

where

R_{pt} is the capitalization weighted placebo cluster return at time t

We run model (3) using full sample data. Due to limitation of Fama-French and Carhart data, we run model (4) using subsample data. The summary statistics are shown in Table 1.

With the presence of placebo cluster returns and lagged placebo cluster returns, the average beta of cluster returns is still positive and significant at 1%. $\bar{\beta}_{Ct}$ in model (3) is 0.58 as shown in Table 4 whereas that in model (4) is 0.30 as shown in Table 5. Though the average placebo cluster beta $\bar{\beta}_{Pt}$ is positive and significant in both models (3) and (4), the magnitude is much lower than $\bar{\beta}_{Ct}$. As shown in Table 4, $\bar{\beta}_{Pt}$ is 0.14 and $\bar{\beta}_{Ct}$is 0.58. As shown in Table 5, $\bar{\beta}_{Pt}$ is 0.05 and $\bar{\beta}_{Ct}$is 0.30. Though individual stock returns may be influenced by factors such as size, value, and industry as captured by the placebo cluster return, the results show that search cluster returns have higher explanatory power in explaining the positive comovement of individual stocks.

Table 4. Comovement Analysis with Cluster Return, Placebo Cluster Return and Market Model

Panel A. Summary Statistics of Betas

	R_{C_it}	R_{C_it-1}	R_{P_it}	R_{P_it-1}	$R_{S\&Pt}$	$R_{S\&Pt-1}$	$News7_{C_it}$	$News14_{C_it}$
Mean	0.58	0.0018	0.14	-0.0055	0.44	0.03	-0.0005	0.0001
SD	0.38	0.21	0.33	0.28	0.59	0.50	0.0098	0.0105
Min	-0.53	-1.05	-1.19	-0.89	-1.97	-2.77	-0.05	-0.06
Max	1.98	0.91	1.66	1.19	2.25	3.08	0.08	0.06

Panel B. Regression Results

	$\bar{R}_{Ct}$	$\bar{R}_{Ct-1}$	$\bar{R}_{Pt}$	$\bar{R}_{Pt-1}$	$\bar{R}_{S\&Pt}$	$\bar{R}_{S\&Pt-1}$	$\overline{News7}_{Ct}$	$\overline{News14}_{Ct}$
Coefficient (Robust SE)	0.58 0.01	0.0018 0.01	0.14 0.01	-0.0055 0.01	0.44 0.03	0.03 0.03	-0.0005 (0.0006)	0.0001 (0.0006)
Bootstrap p-value	0.00	0.41	0.00	0.12	0.00	0.07	0.10	0.33

Table 5. Comovement Analysis with Search Cluster Return, Fama-French and Carhart Factors

Panel A. Summary Statistics of Betas

	R_{C_it}	R_{C_it-1}	R_{P_it}	R_{P_it-1}	$FF49_{it}$	$R_{mt}-R_{ft}$	SMB_t	HML_t	UMD_t	$News7_{C_it}$	$News14_{C_it}$
Mean	0.30	0.01	0.05	0.002	0.79	-0.17	0.12	-0.12	-0.32	-0.0030	-0.0008
SD	0.53	0.26	0.50	0.30	1.35	1.35	0.97	1.23	1.23	0.0220	0.0221
Min	-1.59	-1.13	-1.63	-1.24	-4.53	-10.74	-3.18	-7.66	-9.20	-0.17	-0.20
Max	2.15	1.03	2.10	1.10	10.88	7.71	5.03	6.59	6.05	0.06	0.08

Panel B: Regression Results

	$\bar{R}_{Ct}$	$\bar{R}_{Ct-1}$	$\bar{R}_{Pt}$	$\bar{R}_{Pt-1}$	$\overline{FF49}_t$	$\overline{R_{mt}-R_{ft}}$	$\overline{SMB}_t$	$\overline{HML}_t$	$\overline{UMD}_t$	$\overline{News7}_{Ct}$	$\overline{News14}_{Ct}$
Coefficient (Robust SE)	0.30 (0.02)	0.01 (0.01)	0.05 (0.02)	0.002 (0.01)	0.79 (0.05)	-0.17 (0.06)	0.12 (0.06)	-0.12 (0.06)	-0.32 (0.06)	-0.0030 (0.0010)	-0.0008 (0.0010)
Bootstrap p-value	0.00	0.31	0.01	0.45	0.00	0.05	0.06	0.05	0.00	0.10	0.30

Furthermore, the betas of other control variables are similar to those shown in Tables 2-3. The average beta of market returns $\bar{R}_{S\&Pt}$ in model (3) is 0.44. It is smaller than the average beta of search cluster $\bar{\beta}_{C_t}$, which is 0.58. Also, the average beta of market returns $\overline{R_{mt}-R_{ft}}$ in model (4) is insignificant in the parametric Z test. Similarly, the average betas of lagged returns (e.g. $\bar{\beta}_{Ct-1}$, $\bar{\beta}_{Pt-1}$, $\overline{R_{S\&Pt-1}}$ in model (3) and $\bar{\beta}_{Ct-1}$ and $\bar{\beta}_{Pt-1}$ in model (4)) are all insignificant. Additionally, the average betas of news ($\overline{News7}_{Ct}$ and $\overline{News14}_{Ct}$) are also insignificant in both models. These confirm previous findings that lagged values and news have limited explanatory power in explaining individual stock returns.

In the above analysis, we assume that cluster formation is static over time. However, this assumption is not necessarily true. Over time some stocks may leave a cluster or join a new cluster. As a result, the comovement among those stocks may deteriorate over time due to the changes in user search behaviors. Therefore, our research results may serve as the lowest boundary of actual comovement of stocks in the online habitat. In the subsequent section, we relax the static assumption and analyze the comovement in dynamic clusters. Strong comovement is also observed.

Analysis of Change of Online Habitat

Search clusters exhibit characteristics of habitats. If search clusters do represent investment habitats, then any change in the cluster membership of stocks over time should also impact their return comovement with clusters. More specifically, if a stock enters a new cluster then its return should show greater comovement with the new cluster as compared to the return comovement with its previous (newer) cluster. Likewise, when a stock exits its current cluster, its return comovement will detach from the previous cluster return.

To understand the impact of changes in cluster membership on the comovement, we divided our sample into four time periods, t_1 to t_4. We use the first time period t_1 (5 weeks from 9/15/2011 to 10/19/2011) to determine search clusters. Then in time period t_3 (5 weeks from 12/6/2011 to 1/9/2012), we extracted the search clusters again. Based on the composition of search clusters in different time period, we obtained a list of stocks that changed their cluster membership. We analyzed the comovement of stocks with their clusters in t_2 (10/20/2011 to 12/5/2011) and t_4 (1/10/2012 to 2/24/2012).

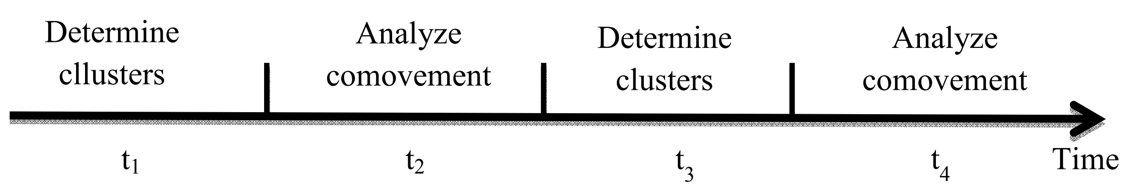

Figure 2. Timeline of comovement analysis

To illustrate the concept of change of online habitat, we use an example as shown in Figure 3. During period t_1, there are three clusters, namely G1 containing stocks R, S, and T; G_2 containing stocks U, V, and W; and G_3 containing stocks X, Y, and Z. Stock Q is not part of any cluster. During period t_3, there are three clusters, namely, H_1 (Q, R, S, T, U), H_2 (V, W), and H_3 (Y, Z) and one independent stock X. From t1 to t_3, we observe three types of changes: addition, deletion and cluster switch. Stock Q, which is independent at t_1, is added to cluster H_1 at t_3. Stock X, which is a member of G_3 in t_1, is deleted from the cluster at time t_3. Finally, stock U switches its cluster membership from G_2 in period t_1 to H_1 in period t_3. In our data sample, we find 22 cluster additions, 80 cluster deletions, and 52 cluster switches. Note than a cluster switch can always be evaluated for addition and deletion. For example, Stock U can also be represented as a deletion from G2 and an addition to H1.

We analyze the change of comovement using modified univariate and bivariate models proposed by Barberis et al. (2005). For additions and deletions, we evaluate the return of stock in periods t_2 and t_4 as a function of the cluster using the following models

$$R_{it_2} = \beta_{0i}^{t_2} + \beta_{1i}^{t_2} R_{S\&Pt_2} + \beta_{2i}^{t_2} R_{S\&Pt_2-1} + \beta_{3i}^{t_2} R_{C_i t_2} + \beta_{4i}^{t_2} R_{C_i t_2-1} + \beta_{5i}^{t_2} \mathrm{News7}_{C_i t_2} + \beta_{6i}^{t_2} \mathrm{News14}_{C_i t_2} + \varepsilon_{it_2} \quad (5)$$

$$R_{it_4} = \beta_{0i}^{t_4} + \beta_{1i}^{t_4} R_{S\&Pt_4} + \beta_{2i}^{t_4} R_{S\&Pt_4-1} + \beta_{3i}^{t_4} R_{C_i t_4} + \beta_{4i}^{t_4} R_{C_i t_4-1} + \beta_{5i}^{t_4} \mathrm{News7}_{C_i t_4} + \beta_{6i}^{t_4} \mathrm{News14}_{C_i t_4} + \varepsilon_{it_4} \quad (6)$$

where

R_{it_s} is the return of stock *i* in period t_s where s = {2, 4}

R_{C_i,t_s} is the return of cluster C associated with stock *i* excluding the return of stock *i*

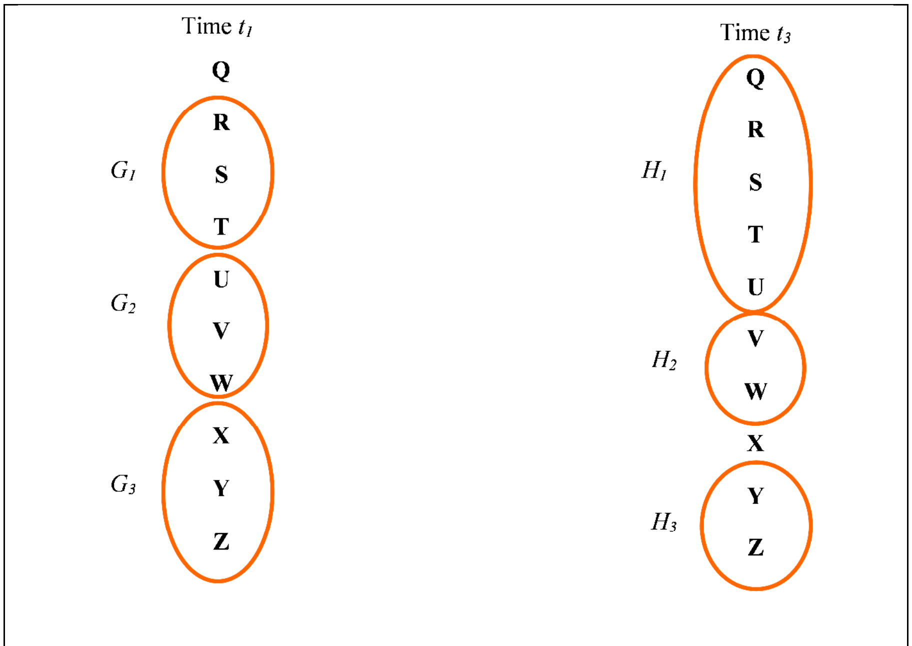

Figure 3. Example of clusters at different time period

A stock could be added to the cluster C in period t_4 or it could be deleted from cluster C in period t_4. According to the theory of comovement, in case of addition, stock i should show higher comovement with cluster C in period t_4 as compared to that in period t_2. Therefore, $\overline{\Delta R}_{Ct}$ should be significant and positive where $\overline{\Delta R}_{Ct}$ is defined as

$$\overline{\Delta R}_{Ct} = \frac{1}{N}\sum_{i=1}^{N}\left(\beta_{3i}^{t_4} - \beta_{3i}^{t_2}\right)$$

where N is the total number of stocks that experience cluster addition or deletion.

Similarly, if a stock is deleted from the cluster, then it should show lower comovement with the cluster in period t_4 as compared to that return in period t_2. In that case $\overline{\Delta R}_{Ct}$ should be significant and negative.

Likewise, following expressions analyze the impact of stocks switching to a different cluster by adapting the bivariate model proposed by Barberis et al. (2005). The return of stock *i* which belongs to cluster G in period t_2 and cluster H in period t_4 can be expressed as

$$R_{it_2} = \beta_{0i}^{t_2} + \beta_{1i}^{t_2}R_{S\&Pt_2} + \beta_{2i}^{t_2}R_{S\&Pt_2-1} + \beta_{3i}^{t_2}R_{G_it_2} + \beta_{4i}^{t_2}R_{G_it_2-1} + \beta_{5i}^{t_2}R_{H_it_2} + \beta_{6i}^{t_2}R_{H_it_2-1} + \beta_{7i}^{t_2}News7_{C_it_2} + \beta_{8i}^{t_2}News14_{C_it_2} + \varepsilon_{it_2} \quad (7)$$

$$R_{it_4} = \beta_{0i}^{t_4} + \beta_{1i}^{t_4}R_{S\&Pt_4} + \beta_{2i}^{t_4}R_{S\&Pt_4-1} + \beta_{3i}^{t_4}R_{G_it_4} + \beta_{4i}^{t_4}R_{G_it_4-1} + \beta_{5i}^{t_4}R_{H_it_4} + \beta_{6i}^{t_4}R_{H_it_4-1} + \beta_{7i}^{t_4}News7_{C_it_4} + \beta_{8i}^{t_4}News14_{C_it_4} + \varepsilon_{it_4} \quad (8)$$

R_{it_s} is the return of stock i in period t_s where s = {2, 4}

R_{G_i,t_s} is the return of cluster G and R_{H_i,t_s} is the return of cluster H

If comovement persists within a cluster then stock *i* should have higher comovement with cluster G relative to cluster H in period t_2, while lower comovement with cluster G as compared to that with cluster H in period t_4. Thus, if *N* is the total number of stocks that experience cluster switch we should expect $\overline{\Delta R}_{Gt}$ to be significant and negative and $\overline{\Delta R}_{Ht}$ be be significant and positive where

$$\overline{\Delta R}_{Gt} = \frac{1}{N}\sum_{i=1}^{N}\left(\beta_{3i}^{t_4} - \beta_{3i}^{t_2}\right) \text{ and } \overline{\Delta R}_{Ht} = \frac{1}{N}\sum_{i=1}^{N}\left(\beta_{5i}^{t_4} - \beta_{5i}^{t_2}\right)$$

As shown in Table 6, in the scenario of cluster addition, $\overline{\Delta R}_{Ct}$ is 0.23, which is positive and significant. This supports our hypothesis that stocks commove more with other stocks in the new habitat after the addition than before the change. However, $\overline{\Delta R}_{Ct-1}$ is found to be insignificant. This is consistent to our prior findings in the previous subsection that lag value has limited power in explaining individual stock comovement. With regard to other controls, they are insignificant in either parametric Z test or block bootstrap test. Though $\overline{\Delta R}_{S\&Pt-1}$ has a significant bootstrap p-value, the parametric Z test statistics is 1.94 (i.e. 0.33/0.17), which is less than the 5% threshold of 1.96.

Table 6. Comovement Analysis of Cluster Addition						
Panel A: Summary Statistics of $\Delta\beta$						
	$\Delta R_{S\&Pt}$	$\Delta R_{S\&Pt-1}$	ΔR_{C_it}	ΔR_{C_it-1}	$\Delta News7_{C_it}$	$\Delta News14_{C_it}$
N=74						
Mean	-0.06	0.33	0.23	-0.04	-0.0070	-0.0049
SD	1.42	1.49	0.89	1.17	0.0900	0.0393
Min	-4.14	-5.55	-2.08	-2.20	-0.66	-0.09
Max	5.13	4.46	3.49	8.23	0.19	0.14
Panel B: Regression Results of Average Betas						
	$\overline{\Delta R}_{S\&Pt}$	$\overline{\Delta R}_{S\&Pt-1}$	$\overline{\Delta R}_{Ct}$	$\overline{\Delta R}_{Ct-1}$	$\overline{\Delta News7}_{Ct}$	$\overline{\Delta News14}_{Ct}$
Coefficient (Robust SE)	-0.06 (0.14)	0.33 (0.17)	0.23 (0.09)	-0.04 (0.11)	-0.0070 (0.0047)	-0.0049 (0.0039)
Bootstrap p-value	0.12	0.01	0.01	0.29	0.25	0.11

With regard to cluster removal, the results show strong support to our hypothesis as shown in Table 7. $\overline{\Delta R}_{Ct}$ is significant and negative with a value of -0.10. The results imply that when a stock moves out of a cluster, its comovement with other stocks in the original cluster drops. The finding is consistent to the theories of comovement. Similar to the findings in the scenario of cluster addition, the lagged variable $\overline{\Delta R}_{Ct-1}$ is insignificant and other control variables are insignificant in either parametric Z test or block bootstrap test.

Table 7. Comovement Analysis of Cluster Removal						
Panel A: Summary Statistics of $\Delta\beta$						
	$\Delta R_{S\&Pt}$	$\Delta R_{S\&Pt-1}$	ΔR_{C_it}	ΔR_{C_it-1}	$\Delta News7_{C_it}$	$\Delta News14_{C_it}$
N=132						
Mean	0.15	0.09	-0.10	-0.01	0.0010	-0.0013
SD	1.39	1.38	0.77	0.57	0.0446	0.0373
Min	-5.36	-4.50	-2.49	-2.41	-0.12	-0.18
Max	4.27	8.59	3.47	2.30	0.25	0.11
Panel B: Regression Results of Average Betas						
	$\overline{\Delta R}_{S\&Pt}$	$\overline{\Delta R}_{S\&Pt-1}$	$\overline{\Delta R}_{Ct}$	$\overline{\Delta R}_{Ct-1}$	$\overline{\Delta News7}_{Ct}$	$\overline{\Delta News14}_{Ct}$
Coefficient (Robust SE)	0.15 (0.11)	0.09 (0.12)	-0.10 (0.05)	-0.01 (0.05)	0.0010 (0.0028)	-0.0013 (0.0027)
Bootstrap p-value	0.03	0.29	0.01	0.63	0.37	0.31

Table 8 shows the results in the scenario of cluster switch. All stocks are members of a cluster at both t_1 and t_3. The results show strong support to our hypothesis. As predicted, $\overline{\Delta R}_{Gt}$ is significant and negative with a value of -0.45 and $\overline{\Delta R}_{Ht}$ is significant and positive with a value of 0.41. The results are consistent to the theories of comovement. When a stock changes its online habitat, it comoves more with stocks in the new cluster than with stocks in the old cluster. Similar to other scenario, the parameters of the lagged terms $\overline{\Delta R}_{Gt-1}$ and $\overline{\Delta R}_{Ht-1}$ are insignificant and other controls are also insignificant in both parametric Z test and block bootstrap test.

Table 8. Comovement Analysis of Cluster Switch

Panel A: Summary Statistics of $\Delta\beta$								
	$\Delta R_{S\&Pt}$	$\Delta R_{S\&Pt-1}$	ΔR_{G_it}	ΔR_{G_it-1}	ΔR_{H_it}	ΔR_{H_it-1}	$\Delta News7_{C_it}$	$\Delta News14_{C_it}$
N=52								
Mean	0.19	0.24	-0.45	-0.08	0.41	-0.06	-0.0010	-0.0038
SD	1.03	1.26	0.52	0.80	0.78	0.67	0.0296	0.0197
Min	-3.64	-1.94	-1.63	-4.14	-1.92	-2.24	-0.13	-0.06
Max	1.95	5.00	0.72	1.83	2.67	1.16	0.05	0.05
Panel B: Regression Results of Average Betas								
	$\overline{\Delta R}_{S\&Pt}$	$\overline{\Delta R}_{S\&Pt-1}$	$\overline{\Delta R}_{Gt}$	$\overline{\Delta R}_{Gt-1}$	$\overline{\Delta R}_{Ht}$	$\overline{\Delta R}_{Ht-1}$	$\overline{\Delta News7}_{Ct}$	$\overline{\Delta News14}_{Ct}$
Coefficient (Robust SE)	0.19 (0.12)	0.24 (0.12)	-0.45 (0.06)	-0.08 (0.06)	0.41 (0.06)	-0.06 (0.07)	-0.0010 (0.0029)	-0.0038 (0.0023)
Bootstrap p-value	0.31	0.06	0.00	0.20	0.00	0.08	0.42	0.21

As a robustness test, we remove all the lagged variables and controls and use the same univariate and bivariate models as suggested by Barberis et al. (2005). The results are qualitatively similar to the results shown above. $\overline{\Delta\beta}$s are significant in all three situations and the sign of $\overline{\Delta\beta}$s is consistent to the theories of comovement. The robustness test results again lend support to our hypothesis. When online habitat alters, the comovement of stocks also changes accordingly.

Robustness Tests

Endogeneity: As returns of stocks within a cluster can influence each other, this can lead to biased estimates of the coefficient for the cluster return. To address the issue, we use message board post volume as an instrumental variable. Previous finance research has shown that the stock messages can influence stock returns (Antweiler and Frank 2004). Message board posts are messages created by investors who are interested in a stock. Most message posters are active retail investors. They gather information from various sources and then share their insights on message boards. These messages in turn can influence the trading activity for a stock. While the posts for a stock can drive its returns, these are not expected to directly influence the returns of another stock but only indirectly through the correlated posts for stocks. Therefore if we also control for the posts for a stock we can use posts for other stocks in the cluster as instrument for the cluster return.

We collect the message post data from Yahoo! Message Board. We use 2SLS to address the endogeneity issue in model (1) First, we run an OLS regression with post volume (*Post*$_{it}$) in model (9). Then we obtain the estimated value of $\widehat{\mathrm{R}}_{\mathrm{jt}}$ and use it to determine capitalization weighted cluster returns $\widehat{\mathrm{R}}_{\mathrm{C_it}}$ as shown in (10). Finally, we use $\hat{R}_{C_it}$ in model (11).

$$\mathrm{R_{it}} = \alpha_{0i} + \alpha_{1i}\mathrm{R}_{S\&Pt} + \alpha_{2i}\mathrm{R}_{S\&Pt-1} + \alpha_{3i}\mathrm{Focal\_News_{it}} + \alpha_{4i}\mathrm{Peer\_News_{C_it}} + \alpha_{5i}\mathrm{Post_{it}} + \varepsilon_{it} \tag{9}$$

where *Post*$_{it}$ is the volume of message board posts of company *i* at time *t* on Yahoo! Message Board

$$\widehat{\mathrm{R}}_{\mathrm{C_it}} = \textstyle\sum_{i\neq j\in C_i} \widehat{\mathrm{R}}_{\mathrm{jt}} \times \mathrm{Cap_j} / \sum_{i\neq j\in C_i} \mathrm{Cap_j} \tag{10}$$

where *Cap*$_j$ is the market capitalization of company *j* in October 2011 (before the time period of comovement analysis)

$$R_{it} = \beta_{0i} + \beta_{1i}\hat{R}_{C_it} + \beta_{2i}\hat{R}_{C_it-1} + \beta_{3i}R_{S\&Pt} + \beta_{4i}R_{S\&Pt-1} + \beta_{5i}Focal\_News_{it} + \beta_{6i}Peer\_News_{C_it} + \beta_{7i}Post_{it} + \varepsilon_{it} \quad (11)$$

The first stage regression of 2SLS as shown in Table 9 shows that message board posts are significant in explaining individual stock returns. The coefficient is positive and significant at 5% in both Z-test and block bootstrapping.

Table 9. First Stage Regression Results of 2SLS with Posts as Instrumental Variable

Panel A: Summary Statistics of $\Delta\alpha$

	$R_{S\&Pt}$	$R_{S\&Pt-1}$	$Focal\_News_{it}$	$Peer\_News_{C_it}$	$Post_{it}$
N=301					
Mean	1.35	0.04	-0.0002	-0.00004	0.0016
SD	0.58	0.25	0.0042	0.0013	0.02
Min	0.16	-1.47	-0.02	-0.01	-0.02
Max	3.23	0.93	0.02	0.01	0.34

Panel B: Regression Results of Average Betas

	$\bar{R}_{S\&Pt}$	$\bar{R}_{S\&Pt-1}$	$\overline{Focal\_News_t}$	$\overline{Peer_{News}}_{Ct-1}$	$\overline{Post_t}$
Coefficient (Robust SE)	1.35 (0.04)	0.04 (0.04	-0.0002 (0.0002)	-0.00004 (0.0001)	0.0016 (0.0002)
Bootstrap p-value	0.00	0.10	0.30	0.42	0.00

In the second stage regression as shown in Table 10, the beta of $\hat{R}_{C_it}$ is positive and significant at 5% in both statistic tests indicating positive comovement.

Table 10. Second Stage Regression Results of 2SLS with Posts as Instrumental Variable

Panel A: Summary Statistics of $\Delta\beta$

	$\hat{R}_{C_it}$	$\hat{R}_{C_it-1}$	$R_{S\&Pt}$	$R_{S\&Pt-1}$	$\overline{Focal\_News_{it}}$	$\overline{Peer_{News}}_{C_it-1}$	$\overline{Post_{it}}$
N=301							
Mean	0.96	-0.09	0.19	0.13	-0.00011	0.00004	0.0016
SD	1.84	1.48	2.32	1.69	0.00440	0.00176	0.02
Min	-5.48	-14.97	-10.12	-5.79	-0.02	-0.01	-0.01
Max	15.89	4.43	14.46	12.87	0.02	0.01	0.33

Panel B: Regression Results of Average Betas

	$\bar{\hat{R}}_{Ct}$	$\bar{\hat{R}}_{Ct-1}$	$\bar{R}_{S\&Pt}$	$\bar{R}_{S\&Pt-1}$	$\overline{Focal\_News_t}$	$\overline{Peer\_News_{Ct}}$	$\overline{Post_t}$
Coefficient (Robust SE)	0.96 (0.12)	-0.09 (0.10)	0.19 (0.16)	0.13 (0.13)	-0.00011 (0.0003)	0.00004 (0.0001)	0.0016 (0.0002)
Bootstrap p-value	0.04	0.26	0.21	0.14	0.32	0.38	0.00

Alternate Placebo Test: In models (3) and (4), we test comovement with placebo cluster return as controls. We try another robustness test by replacing the dependent variable of models (1) and (2) by placebo stock return. The robustness test models are as shown in (12) and (13).

$$R_{P_it} = \beta_{0i} + \beta_{1i}R_{C_it} + \beta_{2i}R_{C_it-1} + \beta_{3i}R_{S\&Pt} + \beta_{4i}Focal\_News_{it} + \beta_{5i}Peer\_News_{C_it} + \varepsilon_{it} \quad (12)$$

$$R_{P_i t} = \beta_{0i} + \beta_{1i} R_{C_i t} + \beta_{2i} R_{C_i t-1} + \beta_{3i} FF49_{it} + \beta_{4i}(R_{mt} - R_{ft}) + \beta_{5i} SMB_t + \beta_{6i} HML_t + \beta_{7i} UMD_t + \beta_{8i} Focal\_News_{it} + \beta_{9i} Peer\_News_{C_i t} + \varepsilon_{it} \quad (13)$$

As shown in Table 11 Panel A, positive comovement is still detected in (12). However, its magnitude, which is 0.16, is much lower than the one reported in (1), which is 0.61. In (13), with additional controls, the positive comovement is no longer detected as shown in Table 11 Panel B. The results show that comovement is strongest among stocks in the same cluster. Other stocks outside the cluster, though similar in industry, size, and value, do not show the same degree of comovement.

Table 11. Robustness Test with Placebo Stock Return as Dependent Variable

Panel (A) Regression Results of Average Betas of Model in CAPM Model (12)

	$\bar{R}_{C_i t}$	$\bar{R}_{C_i t-1}$	$\bar{R}_{S\&Pt}$	$\overline{Focal\_News_{it}}$	$\overline{Peer\_News_{C_i t}}$
Coefficient (Robust SE)	0.16 (0.01)	-0.03 (0.01)	1.02 (0.03)	0.0003 (0.0003)	-0.00004 (0.00005)
Bootstrap p-value	0.00	0.00	0.00	0.15	0.24

Panel (B) Regression Results of Average Betas of Model in Fama-French Model (13)

	$\bar{R}_{Ct}$	$\bar{R}_{Ct-1}$	$\overline{FF49}_t$	$\overline{R_{mt} - R_{ft}}$	$\overline{SMB}_t$	$\overline{HML}_t$	$\overline{UMD}_t$	$\overline{Focal\_News}_t$	$\overline{Peer\_News}_{Ct}$
Coefficient (Robust SE)	-0.03 (0.02)	-0.02 (0.01)	0.01 (0.0005)	0.0018 (0.0005)	0.0031 (0.0004)	0.0010 (0.0004)	-0.0009 (0.0004)	-0.0002 (0.0004)	-0.0001 (0.0002)
Bootstrap p-value	0.12	0.00	0.00	0.03	0.00	0.00	0.03	0.28	0.42

Discussion

We find that stocks in the same search clusters cannot be fully explained by the traditional measures. There can be several alternative explanations. First, investors prefer to purchase stocks with similar risk level. Some risk loving investors may prefer some stocks with high volatility and some risk-averse investors may prefer some stocks with more stable volatility. We determine volatility by regressing individual stock return on S&P 500 40 trading days before the cluster formation time period. The beta of S&P 500 determines the level of volatility. We first compute the volatility ranking of all stocks in CRSP. Then, based on the volatility value, we divide all stocks into 10 deciles. Finally, we compute the volatility similarity index of our sample clusters using formula (14).

$$SI_G = 1 - \left(\frac{\sum_{A \in G} D_A - \bar{D}_G}{4.5} \right)$$

where D_A is the decile of stock A and $\bar{D}_G$ is the mean decile of cluster G and 4.5 is the average possible decile difference (i.e. 0, 1, …, 9). The average volatility similarity of all 72 clusters is 0.80. The results show that many clusters are formed by volatility. For the sake of comparison, we also compute the size and value similarity indices in similar fashion. The average similarity of all 72 clusters is 0.83 for size and 0.71 for value. For the similarity index of industry, we classify all firms based on Fama-French 49 industry classification. Then we find the total number of firms in the same industry and divide it by the total number members in the cluster. The average similarity in industry for all 72 clusters is 0.82.

Apart from volatility, some companies could be co-viewed because they have a strong supplier-customer relationship. Such relationship is not easily captured by traditional size, value, and industry measures. In our sample data, we find that Codexis (Nasdaq: CDXS), which is a bio-catalyst developer, forms a search cluster with its customers Gevo (Nasdaq: GEVO) and Amyris (Nasdaq: AMRS), which are bio-fuel firms. We retrieve supplier-customer relationship from Bloomberg Supply Chain Analysis for all stocks in Russell 3000 index. For each company, we determine how many peers in the same cluster have supplier-customer relationship with it. The maximum number of companies that form a supplier-customer relationship is recorded. The similarity index of supply-chain relationship is to divide the maximum number by the total cluster size. The average supply-chain similarity for all 72 clusters is 0.25. This suggests that not many clusters are formed by supply chain. Nevertheless there are 10 clusters with a similarity index of at least 0.7 indicating that most members within a cluster have supplier-customer relationship.

We test whether similarity in volatility and supplier-customer relationship cause positive comovement. First, we form a placebo stock by matching companies in the clusters with high volatility and supply-chain relationship with another company with data available in CompuStat and CRSP, respectively. We consider only those clusters which have a similarity index of 0.8 or higher for volatility or supply chain. We match each stock in the cluster by choosing a placebo stock that shares the same Fama-French 10 industries and with the smallest difference in market capitalization, price-to-book ratio, and volatility (or with a supply-chain relationship). Then we obtain the regression results using equation (12). Finally, we compare the beta of $\bar{R}_{C_it}$ in (12) with volatility (or supply-chain) placebo stock return as dependent variable and the beta of $\bar{R}_{C_it}$ obtained in (12) with placebo stocks that match only the three fundamentals (size, value, and industry). The beta of $\bar{R}_{C_it}$ with volatility placebo as dependent variable is 0.21 (significant at 1% in robust SE and block bootstrapping). It is higher than the beta obtained with placebo that match only the three fundamentals (size, value, industry), which is 0.165 (significant at 1%) but lower that the beta of $\bar{R}_{C_it}$ for the stocks within the cluster (0.63). The beta obtained with supply-chain placebo as dependent variable is 0.23 (significant at 1%), which is higher than the beta obtained from placebo that match only the three fundamental, which is 0.22 but lower than the beta for stocks within the cluster (0.4). The results provide support to the conjecture that volatility or even supply-chain relationship can lead to additional positive comovement that cannot be explained by the fundamentals. These results also show that the comovement within the search cluster is still higher than these other factors. This suggests that cluster stocks share some other unknown characteristics. It is also possible that some of this comovement is induced by the platform itself.

Conclusion

In this research, we investigate the search behavior of members of Yahoo! Finance for the Russell 3000 stocks. Based on the co-viewing data of stocks we identify several search clusters which represent groups of stocks which investors search together. We find that a positive stock comovement exists among stocks in the same cluster even after controlling for market returns and Fama-French, Carhart factors, and accumulated news in the same cluster. Furthermore, we analyze comovement when stocks experience change in online habitat. We find that the change in comovement is consistent with the theories of stock comovement. When a stock is added to a new search cluster, it comoves more with the stocks in the same cluster after the addition than before. When a stock leaves a cluster, it comoves less with other stocks in the old cluster. When a stock switches from one cluster to another, it comoves more with other stocks in the new cluster and less with those in the old cluster. Using additional tests we establish that the positive comovement is not likely to be purely driven by factors such as size, value, and industry. We show that search clusters can share other characteristics such as similarity in volatility or supply chain which can drive comovement of stocks.

We make several contributions in this paper. First, we add to the existing behavioral finance literature by investigating investment habitats through online user search behaviors. Traditional comovement literature analyzes stock comovement based on styles (e.g. big cap, small cap, value stocks, growth stocks) and market indices (e.g. S&P 500). We find that high frequency of searching can reveal granular habitats. As people view some stocks consistently together, the stocks naturally form a search cluster. We find that comovement is higher for stocks in search clusters as compared to comovement with stocks of similar size, value, and industry. Also, when the search habitat alters over time, the comovement of stocks in the same search clusters also changes. This suggests that search behavior can ex-ante reveal the changes in the comovement of stocks. We also contribute to the IS literature on the online search behavior. We demonstrate how the online search behavior can be used to determine investment habitats and predict their performance.

References

Antweiler, W., and Frank, M.Z. 2004. "Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards," *The Journal of Finance* (59:3), pp. 1259-1294.

Barberis, N., Shleifer, A., and Wurgler, J. 2005. "Comovement," *Journal of Financial Economics* (75:2), pp. 283-317.

Boyer, B.H. 2011. "Style-Related Comovement: Fundamentals or Labels?," *The Journal of Finance* (66:1), pp. 307-332.

Da, Z.H.I., Engelberg, J., and Gao, P. 2011. "In Search of Attention," *The Journal of Finance* (66:5), pp. 1461-1499.

Das, S.R., and Chen, M.Y. 2007. "Yahoo! For Amazon: Sentiment Extraction from Small Talk on the Web," *Mangement Science* (53:9), pp. 1375-1388.

Diether, K.B., Lee, K.-H., and Werner, I.M. 2009. "It's SHO Time! Short-Sale Price Tests and Market Quality," *The Journal of Finance* (64:1), pp. 37-73.

Evans, R.B., Geczy, C.C., Musto, D.K., and Reed, A.V. 2009. "Failure Is an Option: Impediments to Short Selling and Options Prices," *Review of Financial Studies* (22:5), May 1, 2009, pp. 1955-1980.

Froot, K., and Teo, M. 2008. "Style Investing and Institutional Investors," *Journal of Financial & Quantitative Analysis* (43:4), pp. 883-906.

Green, T.C., and Hwang, B.-H. 2009. "Price-Based Return Comovement," *Journal of Financial Economics* (93:1), pp. 37-50.

Greenwood, R. 2008. "Excess Comovement of Stock Returns: Evidence from Cross-Sectional Variation in Nikkei 225 Weights," *Review of Financial Studies* (21:3), May 1, 2008, pp. 1153-1186.

Haugen, R.A., and Baker, N.L. 1996. "Commonality in the Determinants of Expected Stock Returns," *Journal of Financial Economics* (41:3), pp. 401-439.

Kumar, A. 2009. "Who Gambles in the Stock Market?," *The Journal of Finance* (64:4), pp. 1889-1933.

Kumar, A., and Lee, C.M.C. 2006. "Retail Investor Sentiment and Return Comovements," *The Journal of Finance* (61:5), pp. 2451-2486.

Kumar, A., Page, J.K., and Spalt, O.G. 2012. "Investor Sentiment and Return Comovements: Evidence from Stock Splits and Headquarters Changes," *Review of Finance*), April 25, 2012.

Kuruzovich, J., Viswanathan, S., Agarwal, R., Gosain, S., and Weitzman, S. 2008. "Marketspace or Marketplace? Online Information Search and Channel Outcomes in Auto Retailing," *Information Systems Research* (19:2), June 1, 2008, pp. 182-201.

Massa, M., and Zhang, L. 2009. "Cosmetic Mergers: The Effect of Style Investing on the Market for Corporate Control," *Journal of Financial Economics* (93:3), pp. 400-427.

Patton, A., Politis, D., and White, H. 2009. "Correction to "Automatic Block-Length Selection for the Dependent Bootstrap" by D. Politis and H. White," *Econometric Reviews* (28:4), pp. 372-375.

Peng, L., Xiong, W., and Bollerslev, T. 2007. "Investor Attention and Time-Varying Comovements," *European Financial Management* (13:3), pp. 394-422.

Pindyck, R.S., and Rotemberg, J.J. 1993. "The Comovement of Stock Prices," *The Quarterly Journal of Economics* (108:4), November 1, 1993, pp. 1073-1104.

Politis, D.N., and White, H. 2004. "Automatic Block-Length Selection for the Dependent Bootstrap," *Econometric Reviews* (23:1), pp. 53-70.

Vijh, A. 1994. "S&P 500 Trading Strategies and Stock Betas," *Review of Financial Studies* (7:1), January 1, 1994, pp. 215-251.

Wu, L., and Brynjolfsson, E. 2009. "The Future of Prediction: How Google Searches Foreshadow Housing Prices and Sales," *Proceedings of Thirtieth International Conference on Informaiton Systems*, Phoenix, pp. 1-14.