Online Appendix

A. Summary Statistics

Table A.1. Summary Statistics

Panels A and B report the summary statistics for all variables used in the comovement analysis based on baseline (Model 1) and extended Fama-French four-factor (Model 2) models. Panels C and D report the summary statistics for all variables used in the dynamic analysis corresponding to three different scenarios—namely, addition, removal, and switch—in two different time periods t_2 (October to November) and t_4 (January to February).

Panel A: Summary Statistics of Data Used in Model 1

	R_{it}	R_{ct}	R_{ct-1}	$R_{p_i t}$	$R_{S\&Pt}$	$R_{S\&Pt-1}$	$FocalNews_{it}$	$PeerNews_{Cit}$
N (firm-days)	14,490	14,490	14,490	14,490	14,490	14,490	14,490	14,490
Mean	-0.00	-0.00	+0.00	-0.00	-0.00	+0.00	1.81	3.02
SD	0.04	0.03	0.03	0.03	0.02	0.02	1.29	1.75
Min	-0.84	-0.26	-0.26	-0.44	-0.04	-0.04	0.00	0.00
Max	0.67	0.17	0.24	0.51	0.04	0.04	4.47	6.55

Panel B: Summary Statistics of Data Used in Model 2

	R_{it}	R_{ct}	R_{ct-1}	$R_{p_i t}$	R_{mt}-R_{ft}	SMB_t	HML_t	UMD_t	$FocalNews_{it}$	$PeerNews_{Cit}$
N (firm-days)	14,490	14,490	14,490	14,490	14,490	14,490	14,490	14,490	14,490	14,490
Mean	-0.00	-0.00	+0.00	-0.00	-0.00	+0.00	+0.00	+0.00	1.81	3.02
SD	0.04	0.03	0.03	0.03	0.02	0.01	0.01	0.01	1.29	1.75
Min	-0.84	-0.26	-0.26	-0.44	-0.04	-0.01	-0.01	-0.03	0.00	0.00
Max	0.67	0.17	0.24	0.51	0.04	0.02	0.01	0.02	4.47	6.55

Panel C: Summary Statistics at t_2

Scenario	Addition	Deletion	Switch

	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
R_{it_2}	-0.00	0.04	-0.61	0.23	-0.00	0.04	-0.48	0.29	-0.00	0.03	-0.26	0.23
$R_{C_it_2}$	-0.00	0.04	-0.18	0.72	-0.00	0.03	-0.26	0.17	-0.00	0.03	-0.16	0.12
$R_{C_it_2-1}$	-0.00	0.04	-0.18	0.72	+0.00	0.03	-0.26	0.24	-0.00	0.03	-0.16	0.12
$R_{S\&Pt_2}$	-0.00	0.02	-0.04	0.04	-0.00	0.02	-0.04	0.04				
$R_{S\&Pt_2-1}$	+0.00	0.02	-0.04	0.04	+0.00	0.02	-0.04	0.04				
$R_{G_it_2}$									-0.00	0.03	-0.18	0.13
$R_{G_it_2-1}$									-0.00	0.03	-0.18	0.13
$R_{H_it_2}$									-0.00	0.02	-0.04	0.04
$R_{H_it_2-1}$									+0.00	0.02	-0.04	0.04
$MktRf_t$	-0.00	0.02	-0.04	0.04	-0.00	0.02	-0.04	0.04	-0.00	0.02	-0.04	0.04
SMB_t	+0.00	0.01	-0.01	0.02	+0.00	0.01	-0.01	0.02	+0.00	0.01	-0.01	0.02
HML_t	+0.00	0.01	-0.01	0.01	+0.00	0.01	-0.01	0.01	+0.00	0.01	-0.01	0.01
UMD_t	+0.00	0.01	-0.03	0.02	+0.00	0.01	-0.03	0.02	+0.00	0.01	-0.03	0.02
$FocalNews_{it_2}$	1.86	1.33	0.00	4.43	1.78	1.27	0.00	4.47	1.94	1.34	0.00	4.36
$PeerNews_{C_it_2}$	2.22	2.15	0.00	6.32	3.15	1.79	0.00	6.55	3.58	1.73	0.00	6.32

Panel D: Summary Statistics at t_4

Scenario	Addition				Deletion				Switch			
	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
R_{it_4}	+0.00	0.03	-0.30	0.78	+0.00	0.03	-0.28	0.64	+0.00	0.02	-0.08	0.11
$R_{C_it_4}$	+0.00	0.02	-0.13	0.28	+0.00	0.02	-0.19	0.43	+0.00	0.03	-0.15	0.23
$R_{C_it_4-1}$	+0.00	0.02	-0.13	0.28	+0.00	0.02	-0.19	0.43	+0.00	0.03	-0.15	0.23
$R_{S\&Pt_4}$	+0.00	0.01	-0.02	0.02	+0.00	0.01	-0.02	0.02				
$R_{S\&Pt_4-1}$	+0.00	0.01	-0.02	0.02	+0.00	0.01	-0.02	0.02				
$R_{G_it_4}$									+0.00	0.01	-0.04	0.07
$R_{G_it_4-1}$									+0.00	0.01	-0.04	0.07
$R_{H_it_4}$									+0.00	0.01	-0.02	0.02
$R_{H_it_4-1}$									+0.00	0.01	-0.02	0.02
$MktRf_t$	+0.00	0.01	-0.02	0.02	+0.00	0.01	-0.02	0.02	+0.00	0.01	-0.02	0.02
SMB_t	+0.00	0.01	-0.01	0.01	+0.00	0.01	-0.01	0.01	+0.00	0.01	-0.01	0.01

HML_t	-0.00	0.00	-0.01	0.01	-0.00	0.00	-0.01	0.01	-0.00	0.00	-0.01	0.01
UMD_t	-0.00	0.01	-0.02	0.01	-0.00	0.01	-0.02	0.01	-0.00	0.01	-0.02	0.01
$FocalNews_{it_4}$	0.88	0.82	0.00	2.89	0.88	0.81	0.00	3.00	0.90	0.82	0.00	2.89
$PeerNews_{C_i t_4}$	2.31	1.30	0.00	5.10	0.95	1.40	0.00	5.10	2.55	1.35	0.00	5.10

B. Robust Standard Errors of Average Beta

In models (1)-(6), we generate average betas for individual cluster stocks. As all stocks overlap in the same time period, we have to take into consideration the cross-sectional dependence. To address this issue, we compute the robust standard error of average beta $\bar{\beta}$, as the square root of following equation. This approach has been adopted by Boyer (2011).

$$Var(\bar{\beta}) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} Cov(\beta_i, \beta_j)$$

where

n is number of stocks;

Cov(β_i, β_j) is the covariance of beta obtained from stock *i* and that from stock *j*.

Assuming that the residuals of each regression are i.i.d. across time but correlated cross-sectionally, we can estimate the variance covariance matrix across regressions as below:

$$\widehat{\Sigma_{ij}} = \left(X_i^{'}\ X_i\right)^{-1} X_i^{'}\ \left(\frac{\hat{\epsilon}_i\hat{\epsilon}_j{}'}{T}\right) X_j \left(X_j^{'}\ X_j\right)^{-1}$$

where

X_i is the matrix of independent variables of stock *I;*

$\hat{\epsilon}_i$ is the residual of regression of stock *I;*

T is the number of days used in the regression;

Cov(β_i, β_j) where i $\neq$ j is an element of the diagonal matrix $\widehat{\Sigma_{ij}}$.

In order to account for time series correlation within a firm we apply Newey West correction to the SE for each stock.

C. Block Bootstrap Approach in Computing *p*-Value

Consider following regression model:

$$Y_{it} = \beta X_{it} + \varepsilon_{it}$$

where Y is a dependent variable, X is a vector of independent variable, β is a vector of coefficients, ε is a disturbance term, $i = 1, 2, \dots, N$ is panel variable, and $t = 1, 2, \dots, T$ is time variable.

To address the temporal and cross-sectional dependence issues in OLS regression, we adopt a block bootstrap approach.

First, we organize the panel data into overlapping blocks with a temporal block size of L. We follow the approach of Politis and White (2004) and Patton et al. (2009) to find the optimal block size. Patton shares the Matlab code on his website (http://public.econ.duke.edu/~ap172/code.html) and we use the program to find the optimal block size.

Block 1 contains following data:

$\{x_{it}, y_{it}\}$ where $i = 1, 2, \dots, N$ and $t = 1, 2, \dots, L$;

Block 2 contains following data:

$\{x_{it}, y_{it}\}$ where $i = 1, 2, \dots, \mathrm{N}$ and $t = 2, 3, \dots, L+1$;

Block T-L+1 contains following data:

$\{x_{it}, y_{it}\}$ where $i = 1, 2, \dots, \mathrm{N}$ and $t = T-L+1, T-L+2, \dots, T$.

Second, we draw randomly from the above blocks $K = T/L$ times with replacement and form a bootstrap sample.

Third, we run an OLS regression using the bootstrap sample and determine the coefficient estimate $\hat{\beta}^1$.

Fourth, we repeat the sampling and regression estimation procedures in steps two and three 1,000 times to obtain 1,000 bootstrap sample estimates $\{\hat{\beta}^1, \hat{\beta}^2, \dots, \hat{\beta}^{1,000}\}$.

Finally, we determine the *p*-value by computing the proportion of bootstrap sample estimates $\hat{\beta}^j < 0$, where $j = 1, 2, \ldots, 1{,}000$ if the coefficient of estimated coefficient (β) is positive. If β is negative, we compute the *p*-value by finding the proportion of bootstrap sample estimates $\hat{\beta}^j > 0$.

D. Extended Comovement Analysis

Table D.1. Results of Extended Comovement Analysis

Panel A: Original Search Clusters Defined in September 2011

New Cluster definition time period	Model	Recent Cluster Return $\bar{\boldsymbol{R}}_{\boldsymbol{Ct}}$		Original Cluster Return $\bar{\boldsymbol{R}}_{\boldsymbol{Ct}}$		Difference between Recent Cluster Return and Original Cluster Return $\Delta\boldsymbol{R}_{\boldsymbol{Ct}}$	
		Coefficient	*p-value*	*Coefficient*	*p-value*	*Coefficient*	*p-value*
Oct 2011	A	0.69 (0.04)	0.00	0.22 (0.06)	0.00	0.47 (0.07)	0.00
	B	0.57 (0.05)	0.00	0.07 (0.05)	0.08	0.50 (0.07)	0.00
Nov 2011	A	0.62 (0.05)	0.00	0.37 (0.05)	0.00	0.25 (0.07)	0.00
	B	0.53 (0.06)	0.00	0.33 (0.05)	0.00	0.20 (0.07)	0.00
Dec 2011	A	0.50 (0.09)	0.00	0.23 (0.03)	0.00	0.27 (0.09)	0.00
	B	0.48 (0.09)	0.00	0.18 (0.03)	0.00	0.30 (0.09)	0.00
Jan 2012	A	0.52 (0.06)	0.00	0.29 (0.05)	0.00	0.23 (0.07)	0.00
	B	0.36 (0.06)	0.00	0.24 (0.04)	0.00	0.13 (0.07)	0.00
Feb 2012	A	0.44 (0.06)	0.00	0.33 (0.06)	0.00	0.11 (0.08)	0.15
	B	0.28 (0.06)	0.00	0.23 (0.06)	0.00	0.05 (0.08)	0.09
Mar 2012	A	0.49 (0.07)	0.00	0.52 (0.08)	0.00	-0.03 (0.10)	0.52
	B	0.36 (0.08)	0.00	0.41 (0.08)	0.00	-0.05 (0.11)	0.45
Apr 2012	A	0.75 (0.04)	0.00	0.28 (0.05)	0.00	0.46 (0.06)	0.00
	B	0.68 (0.05)	0.00	0.15 (0.05)	0.01	0.53 (0.07)	0.00
May 2012	A	0.65 (0.06)	0.00	0.23 (0.06)	0.00	0.42 (0.08)	0.00
	B	0.60 (0.07)	0.00	0.06 (0.07)	0.21	0.54 (0.10)	0.00
Jun 2012	A	0.77 (0.07)	0.00	0.27 (0.05)	0.00	0.51 (0.08)	0.00
	B	0.60 (0.08)	0.00	0.20 (0.06)	0.01	0.40 (0.09)	0.00
Jul 2012	A	0.73 (0.06)	0.00	0.40 (0.06)	0.00	0.33 (0.07)	0.00
	B	0.53 (0.06)	0.00	0.24 (0.06)	0.00	0.29 (0.08)	0.00
Aug 2012	A	0.72 (0.06)	0.00	0.31 (0.05)	0.00	0.41 (0.07)	0.00
	B	0.59 (0.07)	0.00	0.21 (0.05)	0.00	0.38 (0.08)	0.01
Sep 2012	A	0.63 (0.07)	0.00	0.26 (0.04)	0.00	0.37 (0.07)	0.00
	B	0.49 (0.07)	0.00	0.17 (0.04)	0.00	0.32 (0.08)	0.00
Oct 2012	A	0.57 (0.06)	0.01	0.30 (0.05)	0.00	0.27 (0.07)	0.04
	B	0.51 (0.07)	0.00	0.26 (0.05)	0.00	0.25 (0.07)	0.06
Nov 2012	A	0.54 (0.07)	0.00	0.26 (0.05)	0.00	0.28 (0.08)	0.00
	B	0.46 (0.07)	0.00	0.20 (0.05)	0.00	0.25 (0.08)	0.00
Dec 2012	A	0.54 (0.09)	0.00	0.29 (0.04)	0.00	0.25 (0.10)	0.00
	B	0.46 (0.08)	0.00	0.21 (0.04)	0.00	0.25 (0.09)	0.00

Panel B: Original Search Clusters Defined in January 2012

New Cluster definition time period	Model	Recent Cluster Return $\bar{\boldsymbol{R}}_{\boldsymbol{Ct}}$	Original Cluster Return $\bar{\boldsymbol{R}}_{\boldsymbol{Ct}}$	Difference between Recent Cluster Return and Original Cluster Return $\Delta\boldsymbol{R}_{\boldsymbol{Ct}}$

		Coefficient	p-value	Coefficient	p-value	Coefficient	p-value
Feb 2012				No Cluster Switch			
Mar 2012				No Cluster Switch			
Apr 2012	A	0.70 (0.04)	0.00	-0.08 (0.06)	0.06	0.78 (0.08)	0.00
	B	0.64 (0.05)	0.00	-0.04 (0.06)	0.07	0.68 (0.07)	0.00
May 2012	A	0.66 (0.07)	0.00	-0.05 (0.07)	0.29	0.70 (0.10)	0.00
	B	0.62 (0.08)	0.00	-0.08 (0.08)	0.21	0.70 (0.12)	0.00
Jun 2012	A	0.82 (0.07)	0.00	0.17 (0.09)	0.02	0.65 (0.12)	0.00
	B	0.63 (0.08)	0.00	0.19 (0.09)	0.03	0.44 (0.12)	0.01
Jul 2012	A	0.73 (0.06)	0.00	0.21 (0.09)	0.04	0.52 (0.12)	0.00
	B	0.53 (0.07)	0.00	0.31 (0.10)	0.10	0.22 (0.12)	0.06
Aug 2012	A	0.74 (0.07)	0.00	0.08 (0.08)	0.09	0.66 (0.11)	0.00
	B	0.59 (0.08)	0.00	0.08 (0.08)	0.08	0.52 (0.12)	0.01
Sep 2012	A	0.65 (0.07)	0.00	-0.21 (0.09)	0.00	0.86 (0.12)	0.00
	B	0.51 (0.07)	0.00	-0.06 (0.08)	0.16	0.57 (0.11)	0.00
Oct 2012	A	0.59 (0.06)	0.01	-0.04 (0.14)	0.36	0.63 (0.15)	0.02
	B	0.53 (0.07)	0.01	-0.08 (0.11)	0.20	0.60 (0.14)	0.03
Nov 2012	A	0.57 (0.07)	0.00	0.02 (0.09)	0.39	0.55 (0.12)	0.00
	B	0.49 (0.07)	0.00	-0.08 (0.08)	0.24	0.57 (0.12)	0.01
Dec 2012	A	0.57 (0.10)	0.00	0.01 (0.09)	0.42	0.57 (0.14)	0.00
	B	0.52 (0.09)	0.00	-0.06 (0.07)	0.32	0.58 (0.12)	0.00

Panel C: Original Search Clusters Defined in May 2012

New Cluster definition time period	Model	Recent Cluster Return $\bar{\boldsymbol{R}}_{\boldsymbol{Ct}}$		Original Cluster Return $\bar{\boldsymbol{R}}_{\boldsymbol{Ct}}$		Difference between Recent Cluster Return and Original Cluster Return $\Delta\boldsymbol{R}_{\boldsymbol{Ct}}$	
		Coefficient	*p-value*	*Coefficient*	*p-value*	*Coefficient*	*p-value*
Jun 2012	A	0.73 (0.06)	0.00	0.07 (0.06)	0.17	0.67 (0.08)	0.00
	B	0.58 (0.07)	0.00	-0.04 (0.06)	0.18	0.63 (0.09)	0.00
Jul 2012	A	0.72 (0.05)	0.00	0.08 (0.07)	0.06	0.64 (0.08)	0.00
	B	0.55 (0.06)	0.00	-0.10 (0.07)	0.11	0.65 (0.09)	0.00
Aug 2012	A	0.72 (0.06)	0.00	0.08 (0.07)	0.13	0.64 (0.08)	0.00
	B	0.62 (0.07)	0.00	0.01 (0.07)	0.40	0.61 (0.10)	0.00
Sep 2012	A	0.60 (0.06)	0.00	-0.05 (0.06)	0.20	0.64 (0.09)	0.00
	B	0.51 (0.07)	0.00	-0.07 (0.06)	0.18	0.58 (0.09)	0.00
Oct 2012	A	0.50 (0.06)	0.02	-0.19 (0.07)	0.01	0.69 (0.09)	0.00
	B	0.44 (0.06)	0.01	-0.20 (0.07)	0.01	0.64 (0.10)	0.00
Nov 2012	A	0.46 (0.06)	0.00	0.00 (0.08)	0.51	0.46 (0.11)	0.00
	B	0.40 (0.06)	0.00	-0.02 (0.08)	0.15	0.42 (0.11)	0.00
Dec 2012	A	0.46 (0.08)	0.00	0.18 (0.07)	0.03	0.29 (0.11)	0.00
	B	0.41 (0.08)	0.00	0.02 (0.08)	0.53	0.39 (0.11)	0.00

E. Example for Computing Search Cluster Similarity

Let G = {*A*, *B*, *C*, …, *N*} be a search cluster with *n* stocks from A to N.

dim = {*s*, *v*, *i*} where *s*, *v*, and *i* represent the dimension of size, value, and industry, respectively.

s = {1, 2, …, 10} where 1 is the lowest decile and 10 is the highest decile in market capitalization;

v = {1, 2, …, 10} where 1 is the lowest decile and 10 is the highest decile in price-to-book ratio;

i = {1, 2, …, 10} where the number correspond to Fama-French 10 industries.

We obtain similarity index (SI_G) in *s* and *v* for Group G using following formula:

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

where D_A is the decile of stock A in s or *v*, $\overline{D}_G$ is the average decile of stocks in group G in *s* or *v*, $|D_A - \overline{D}_G|$ is the absolute value of the difference between the decile of stock A and the average decile of group G, 4.5 is the normalization factor which is the average difference of all possible deciles, and N_G is the total number stocks in group G.

When we analyze volatility, we classify all stocks in CRSP with valid stock data into 10 deciles. Then we compute the similarity index in volatility using the above formula.

We compute the similarity index in *i* using the following formula:

$$SI_G = \frac{\sum_{A \in G} I(D_A = mode(D_G))}{N_G},$$

where $mode(D_G)$ is the mode of decile of all stocks in group G, $I(D_A = mode(D_G)$ is an indicator function, which is 1 if $D_A = mode(D_G)$ is true and 0 otherwise, and N_G is the total number stocks in group *G*. If there is no mode, SI_G is 0.

We compute the similarity index in supply-chain as follows:

$$SI_G = \frac{N_{SC}}{N_G}$$

$$N_{SC} = \begin{cases} 0 \; if \; \max_{A \in G} \sum_{B \neq A, B \in G} I(S_A = S_B) = 0 \\ \max(\max_{A \in G} \sum_{B \neq A, B \in G} I(S_A = S_B), 2) \end{cases}$$

where N_{SC} is the number of stocks in G with supply-chain relationship, N_G is the total number stocks in group G, $S_A = S_B$ is a logical test of whether stock A and stock B have a supply chain relationship, and $I(S_A = S_B)$ is an indicator function, which is 1 if stock A and stock B have a supply chain relationship and 0 otherwise.

Consider a cluster that contains stocks ARG, PX, and APD with following values in the 3-dimension space:

Ticker	Size (Decile)	Value (Decile)	Fama-French 10-Industry ID	Volatility (Decile)	Number of supply-chain relationship with the peers
ARG	9	8	7	5	1 (with PX)
PX	10	9	3	4	1 (with ARG)
APD	10	8	3	5	0

Using the above formulae, we obtain following similarity index (SI) for each pair of stocks

SI_G in s = $1 - \left(\frac{|9-9.67|}{4.5} + \left|\frac{10-9.67}{4.5}\right| + \left|\frac{10-9.67}{4.5}\right|\right) \Big/ 3 = 0.90;$

SI_G in v = $1 - \left(\left|\frac{8-8.33}{4.5}\right| + \left|\frac{9-8.33}{4.5}\right| + \left|\frac{8-8.33}{4.5}\right|\right) \Big/ 3 = 0.90;$

SI_G in i = $\left(I(D_{ARG} = 3) + I(D_{PX} = 3) + I(D_{APD} = 3)\right) \Big/ 3 = (0 + 1 + 1) \big/ 3 = 0.67;$

SI_G in *volatility* = $1 - \left(\left|\frac{5-4.67}{4.5}\right| + \left|\frac{4-4.67}{4.5}\right| + \left|\frac{5-4.67}{4.5}\right|\right) \Big/ 3 = 0.90;$

SI_G in *supply-chain* = $2 \big/ 3 = 0.67.$

F. Volatility and Supply Chain Similarity

As users search for stocks based on their interests, the search behavior should reveal other characteristics that are common across stocks in a search cluster. These common characteristics could be another source of comovement among cluster stocks. We consider two such characteristics: volatility and supply-chain relationship.

Volatility: Investors have different risk preferences. Previous research finds that individual risk attitude has direct relationship to market participation (Fellner and Maciejovsky 2007). In behavioral finance research, it is found that less sophisticated investors tend to purchase stocks with high volatility or high market risk (Kumar 2009). Some socioeconomic factors—for example, income, education level, occupation, ethnicity, and religion—may also help explain investors' preference for risky stocks (Kumar 2009, Kumar et al. 2011). Thus, it is possible that investors search for stocks based on their volatility or risk and trade systematically, leading to a comovement between such stocks. Therefore, we posit that some search clusters are formed by personal preference of risk level, which is determined by volatility.

To compute volatility, we use 40 daily trading data of each stock in CRSP before the period of cluster identification (i.e., September 2011) and estimate the beta, which is the volatility, in the model:

$$R_{it} = \alpha_i + \beta S\&P_t + \varepsilon_{it}$$

Then we sort all stocks with data available in CRSP and determine their deciles. We determine the similarity in volatility for every cluster using the same approach as the one we adopted for determining size and value similarity. As shown in Table 4, the average cluster similarity index in volatility is 0.77, which is higher than that in value whose average cluster similarity index is 0.72. The results suggest that many stocks in many search clusters are similar in terms of their volatility or risk.

Supply Chain: Firms that are in the same supply chain are also likely to influence each other through the supplier and customer relationship. Prior study finds that there exists a direct relationship between buyers' forecasting behaviors and supplier's delivery performance (Terwiesch et al. 2005). Shocks related to the suppliers (for example, supply shortage and price increase) may exert pressure on customers' profit margins. Similarly, if customers go bankrupt, it may influence the accounts receivables of suppliers.

Therefore, the companies within the same supply chain network may be interrelated financially to each other. Improvements to the supply chain lead to improvements in the financial performance of firms (Dehning et al. 2007). Prior research finds that stocks in tightly connected supplier and customer industries can cross-predict the returns of each other due to diffusion of value-relevant information (Menzly and Ozbas 2010). As firms with strong supply chain relationships are interdependent, it is possible that investors are likely to search related information for both companies. Therefore, strong supplier-customer relationships may help explain the significant comovement.

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. These firms are not in the same industry but they are part of the same supply chain. According to Fama-French industry classification, CDXS (SIC: 2836) belongs to Healthcare, Medical Equipment and Drugs whereas GEVO (SIC: 2860) and AMRS (SIC: 2860) belong to Manufacturing–Machinery, Trucks, Planes, Chemicals, Office Furniture, Paper, Computer Printing.

We determine the similarity of stocks in a search cluster based on their supply chain membership. We collect supply chain relationship data of all members of Russell 3000 using Bloomberg Supply-Chain Analysis. We construct a supply chain similarity index using the formulae below:

$$SI_G = \frac{N_{SC}}{N_G} \text{ and } N_{SC} = \begin{cases} 0 \; if \;\; \max_{A \in G} \sum_{B \neq A, B \in G} I(S_A = S_B) = 0 \\ \max(\max_{A \in G} \sum_{B \neq A, B \in G} I(S_A = S_B), 2) \end{cases}$$

N_{SC} is the number of stocks in G with supply-chain relationship, N_G is the total number stocks in cluster G, $S_A = S_B$ is a logical test of whether stock A and stock B have a supply chain relationship, and $I(S_A = S_B)$ is an indicator function, which is 1 if stock A and stock B have a supply chain relationship and 0 otherwise. Stocks in a cluster can belong to different supply chains. It is also possible that a stock may be associated with two different supply chains in a cluster. We tag a firm with the dominant supply chain and determine the fraction of firms in the cluster associated with the dominant supply chain. Appendix E shows an example in computing the similarity index. As shown in Table 4, the average cluster similarity index in a supply chain is only 0.24. This suggests that fewer clusters share similarity in a supply chain.

Comovement Comparison: We test whether volatility and supply chain relationship contribute to the incremental comovement separately among clusters with high similarity in volatility and supply chain relationship. We consider a cluster to be highly similar in a characteristic if the similarity index of the cluster for that characteristic is above the third quartile among all identified clusters. To investigate the incremental contribution of volatility (supply chain), we first find a placebo stock that matches closely to a cluster stock in size, value, and industry. Second we find volatility (supply chain) placebo stock that matches individual cluster stock in size, value, industry, and volatility (supply chain). For volatility placebo, we consider stocks in the same decile for size, industry, value and volatility as the cluster stock. Next, we rank all potential placebo stocks independently according to the absolute differences in the value of these characteristics (higher rank for lower difference). Then, we sum up the ranks and pick the matching stock with the highest rank (lowest value). For the supply chain placebo, we consider stocks that match with the cluster stocks in terms of deciles for size, industry, and value. Among the potential placebo stocks, we pick the one that has a supply chain relationship with the cluster stock. We estimate our main model (1) using these placebo stocks to determine the comovement of these placebo stocks with the search cluster. We also re-estimate the main model using the return of cluster stocks as the dependent variable for these clusters with high similarity in volatility (supply chain). Tables F.1 and F.2 summarize the results for the coefficients of the cluster returns for the three different types of stocks.

Table E.1 (Table E.2) consistently shows that the comovement of cluster stock returns is significant and higher than the comovement of the placebo stocks and as well as the volatility (supply chain) placebo stock. The differences of average betas are all significant and positive at 1%. We also find that the comovement of the supply chain and volatility placebo stocks is higher than that of the placebo stocks based on matching of size, industry, and value. These results clearly show that volatility (supply chain relationship) increases the magnitude of comovement and can explain the higher comovement for some search clusters where the similarity is high for these attributes. However, the fact the overall cluster comovement is the strongest suggests that the cluster stocks may possess some additional characteristics that can lead to higher comovement as compared to different placebo stocks.

As supply chain stocks may reside in the same industry, it is possible that the results for the comovement of supply chain stocks are primarily driven by a better match in the industry as compared to that obtained by using Fama-French 10 industry classification. In order to validate that is not the case, we conduct another robustness test by finding placebo stocks that match in size and value but not in the same Fama-French 10 industry. As for the supply chain placebo, we find stocks that match in size and value and pick the one with highest supply chain relationship value. The relationship value is obtained from Bloomberg's supply chain data, which defines the value to be total monetary amount between two companies in the supply chain relationship. As some companies do not have relationship data, we are unable to find corresponding supply chain placebo stock and thus they are removed from the analysis. The comparison results are as shown in Table E.3.

The coefficient for the cluster stock is positive and significant (Table E.3). However, the coefficients for the placebo based on size and value match and supply chain placebo are smaller in magnitude. We again find that the coefficient has the highest value for the cluster stock. The difference of average betas between cluster stocks and supply chain stocks and between cluster stocks and placebo stocks are all significant and positive at 1%. The results lend further support to the conjecture that strong supply chain relationships lead to significant comovement in some search clusters.

Table F.1. Comovement Comparison of Cluster Stock, Placebo Stock, and Volatility Placebo

Panel A: Summary Statistics of Estimated Betas in Model 1

Size = 341	R_{C_it} for Cluster Stock (1)	R_{C_it} for Placebo Stock (2)	R_{C_it} for Volatility Placebo Stock (3)	ΔR_{C_it} (1) – (2)	ΔR_{C_it} (1) – (3)	ΔR_{C_it} (2) – (3)
Mean	0.57	0.14	0.19	0.43	0.38	-0.05
(Robust SE)	(0.03)	(0.04)	(0.03)	(0.05)	(0.04)	(0.03)
Bootstrap p-value	0.00	0.00	0.00	0.00	0.00	0.03

Panel B: Summary Statistics of Estimated Betas in Model 2

Size = 341	R_{C_it} for Cluster Stock (1)	R_{C_it} for Placebo Stock (2)	R_{C_it} for Volatility Placebo Stock (3)	ΔR_{C_it} (1) – (2)	ΔR_{C_it} (1) – (3)	ΔR_{C_it} (2) – (3)

Mean	0.48	0.05	0.08	0.43	0.39	-0.04
(Robust SE)	(0.03)	(0.04)	(0.04)	(0.05)	(0.05)	(0.03)
Bootstrap *p*-value	0.00	0.05	0.00	0.00	0.00	0.05

Table F.2 Comovement Comparison of Cluster Stock, Placebo Stock, and Supply Chain Placebo

Panel A: Summary Statistics of Estimated Betas in Model 1

Size = 97	R_{C_it} for Cluster Stock (1)	R_{C_it} for Placebo Stock (2)	R_{C_it} for Supply Chain Placebo Stock (3)	ΔR_{C_it} (1) – (2)	ΔR_{C_it} (1) – (3)	ΔR_{C_it} (2) – (3)
Mean	0.57	0.19	0.22	0.39	0.35	-0.03
(Robust SE)	(0.04)	(0.05)	(0.05)	(0.06)	(0.06)	(0.04)
Bootstrap *p*-value	0.00	0.00	0.00	0.00	0.00	0.27

Panel B: Summary Statistics of Estimated Betas in Model 2

Size = 97	R_{C_it} for Cluster Stock (1)	R_{C_it} for Placebo Stock (2)	R_{C_it} for Supply Chain Placebo Stock (3)	ΔR_{C_it} (1) – (2)	ΔR_{C_it} (1) – (3)	ΔR_{C_it} (2) – (3)
Mean	0.40	0.03	0.06	0.37	0.35	-0.02
(Robust SE)	(0.05)	(0.06)	(0.06)	(0.07)	(0.07)	(0.05)
Bootstrap *p*-value	0.00	0.29	0.13	0.00	0.00	0.43

Table F.3 Comovement Comparison of Cluster Stock, Placebo Stock, and Supply Chain Placebo with Relationship Value

Panel A: Summary Statistics of Estimated Betas in Model 1

Size = 82	R_{C_it} for Cluster Stock (1)	R_{C_it} for Placebo Stock (2)	R_{C_it} for Supply Chain Placebo Stock (3)	ΔR_{C_it} (1) – (2)	ΔR_{C_it} (1) – (3)	ΔR_{C_it} (2) – (3)
Mean	0.60	0.18	0.14	0.41	0.45	0.04
(Robust SE)	(0.05)	(0.05)	(0.05)	(0.06)	(0.07)	(0.06)
Bootstrap *p*-value	0.00	0.00	0.02	0.00	0.00	0.28

Panel B: Summary Statistics of Estimated Betas in Model 2

Size = 82	R_{C_it} for Cluster Stock (1)	R_{C_it} for Placebo Stock (2)	R_{C_it} for Supply Chain Placebo Stock (3)	ΔR_{C_it} (1) – (2)	ΔR_{C_it} (1) – (3)	ΔR_{C_it} (2) – (3)
Mean	0.44	0.08	0.13	0.36	0.31	-0.06
(Robust SE)	(0.05)	(0.06)	(0.07)	(0.08)	(0.09)	(0.08)
Bootstrap *p*-value	0.00	0.06	0.01	0.00	0.00	0.26

G. Endogeneity Due to Co-Viewing

Investors may click on co-viewed stocks presented to them. We use an instrumental variable approach to account for the potential endogeneity and then compare the consistency of the IV and the OLS estimates. Prior research finds that online investor sentiments expressed in message boards have a statistically significant effect on stock returns (Antweiler and Frank 2004, Sabherwal et al. 2011). We use Yahoo! Message board posts to determine the sentiments. People who post sentiments are more sophisticated investors with subjective opinions on particular stocks and they are not likely to be herded or influenced by the also-viewed list on Yahoo! Finance. Therefore, our sentiment index is exogenous from the potential "herding" effect induced by the also-viewed list. Further, if we control for the sentiments of a stock then we can expect sentiments for other stocks to not influence the returns for a stock. Therefore, if we control for the sentiment for a stock, we can use sentiments for other stocks in the cluster as instrument for the cluster return.

We use Yahoo! message boards, to extract trading sentiments. We adopt Antweiler and Frank's (2004) bullishness formula to measure investor sentiment. $Sentiment_{it}$ of a stock i at time t can be expressed as

$$Sentiment_{it} = B_t \ln(1 + M_t)$$

where B_t is bullishness score and $\ln(1 + M_t)$ is a weight associated with the bullish score. M_t is the number of messages with different sentiments. The higher the number of individuals expressing their sentiments, the higher the weighting. B_t is defined as

$$B_t = \frac{M_t^{Buy} - M_t^{Sell}}{M_t^{Buy} + M_t^{Buy}}$$

where M_t^{Buy} (M_t^{Sell}) is the total number of bullish (bearish) messages on day t. Following Antweiler and Frank (2004) and Sabherwal et al. (2011), we use current sentiments as well as sentiments with one-day and two-day lags as instrumental variables.

We use a 2SLS approach to correct for the potential endogeneity. In the first stage, we run an OLS regression on individual stocks as below:

$$R_{it} = \alpha_0 + \alpha_1 Sentiment_{it} + \alpha_2 \overline{Sentiment}_{it-1,\mathrm{t}-2} + \alpha_3 R_{\mathrm{S\&P}t} + \alpha_4 R_{\mathrm{S\&P}t-1} + \alpha_5 FocalNews_{it} + \alpha_6 PeerNews_{C_i t} + \varepsilon_{it} \quad \text{(i)}$$

$$R_{it} = \alpha_0 + \alpha_1 Sentiment_{it} + \alpha_2 \overline{Sentiment}_{it-1,\mathrm{t}-2} + \alpha_3 MktRf_t + \alpha_4 SMB_t + \alpha_5 HML_t + \alpha_6 UMD_t + \alpha_7 FocalNews_{it} + \alpha_8 PeerNews_{C_i t} + \varepsilon_{it} \quad \text{(ii)}$$

We then compute estimated cluster return using the formula below.

$$\hat{R}_{C_i t} = \sum_{i \neq j \in C_i} \hat{R}_{jt} \times Cap_j \Big/ \sum_{i \neq j \in C_i} Cap_j$$

Finally, we plug in the predicted value of, $\hat{R}_{C_i t-1}$, $\hat{R}_{C_i t}$, in the second stage as shown in (iii) and (iv).

$$R_{it} = \beta_0 + \beta_1 \hat{R}_{C_i t} + \beta_2 \hat{R}_{C_i t-1} + \beta_3 Sentiment_{it} + \beta_4 \overline{Sentiment}_{it-1,\mathrm{t}-2} + \beta_5 R_{S\&Pt} + \beta_6 R_{S\&Pt-1} + \beta_8 FocalNews_{it} + \beta_9 PeerNews_{C_i t} + \varepsilon_{it} \quad \text{(iii)}$$

$$R_{it} = \beta_0 + \beta_1 \hat{R}_{C_i t} + \beta_2 \hat{R}_{C_i t-1} + \beta_3 Sentiment_{it} + \beta_4 \overline{Sentiment}_{it-1,\mathrm{t}-2} + \alpha_5 MktRf_t + \alpha_6 SMB_t + \alpha_7 HML_t + \alpha_8 UMD_t + \beta_9 FocalNews_{it} + \beta_{10} PeerNews_{C_i t} + \varepsilon_{it} \quad \text{(iv)}$$

We performed an F-test in the first stage for each of the instruments. In each case, the F-test value was well over 10, suggesting that our instruments are not weak. In addition, the Hansen's J-Test could not reject the null hypothesis of valid over-identifying restrictions.

The regression results are as shown in Tables F.1 and F.2. As shown in Table F.1, Panels A and B, the coefficients of the instruments, $Sentiment_{it}$ and $\overline{Sentiment}_{it-1,\mathrm{t}-2}$, are both positive and significant in equations (i) and (ii). The market return $R_{\mathrm{S\&P}t}$ is significant in Model (i) and the 4 risk factors in Model (ii) are all significant. The news factors are insignificant in both models.

The results in the second stage model are consistent with our main results. The predicted cluster return $\hat{R}_{C_i t}$ is significant and positive while the lagged cluster return $\hat{R}_{C_i t-1}$ is insignificant. We use the Hausman specification test to validate the null hypothesis that both OLS and 2SLS estimates are consistent (Greene 2003). The Wald *t* statistics are 5.66 and 3.88 for the second stage equations (iii) and (iv). For a Chi square degrees of freedom of 2, both test statistics have *p*-value greater than 0.05. This suggests that the difference between the 2SLS estimates and OLS estimates is not significant. These results provide evidence that Yahoo! Finance does not induce comovement between stocks and the results obtained from our original model are not biased.

Table G.1. Results for First Stage Regression

Panel A: Summary Statistics of Estimates in Model (i)

	$Sentiment_{it}$	$\overline{Sentiment}_{it-1,t-2}$	$R_{S\&Pt}$	$R_{S\&Pt-1}$	$FocalNews_{it}$	$PeerNews_{C_it}$
Coeff	0.0045	0.0037	1.35	0.02	0.0002	-0.0007
(Robust SE)	(0.0012)	(0.0016)	(0.04)	(0.04)	(0.0007)	(0.0005)
Bootstrap *p*-value	0.00	0.03	0.00	0.13	0.38	0.22

Panel B: Summary Statistics of Estimates in Model (ii)

	$Sentiment_{it}$	$\overline{Sentiment}_{it-1,t-2}$	$MktRf_t$	SMB_t	HML_t	UMD_t	$FocalNews_{it}$	$PeerNews_{C_it}$
Coeff	0.0036	0.0029	0.85	0.46	-0.27	-0.66	0.0004	-0.0002
(Robust SE)	(0.0011)	(0.0013)	(0.06)	(0.09)	(0.10)	(0.10)	0.0006	0.0005
Bootstrap *p*-value	0.01	0.05	0.00	0.00	0.01	0.00	0.34	0.57

Table G.2. Results of Second Stage Regression

Panel A: Summary Statistics of Estimates in Model (iii)

	$\hat{R}_{C_i t}$	$\hat{R}_{C_i t-1}$	$Sentiment_{it}$	$\overline{Sentiment}_{it-1,t-2}$	$R_{S\&Pt}$	$R_{S\&Pt-1}$	$FocalNews_{it}$	$PeerNews_{C_i t}$
Coeff	0.79	-0.22	0.0047	0.0044	0.20	0.29	0.0001	-0.0002
(Robust SE)	(0.13)	(0.12)	(0.0011)	(0.0016)	(0.16)	(0.14)	(0.0007)	(0.0006)
Bootstrap *p*-value	0.00	0.03	0.01	0.02	0.04	0.03	0.42	0.33

Panel B: Summary Statistics of Estimates in Model (iv)

	$\hat{R}_{C_i t}$	$\hat{R}_{C_i t-1}$	$Sentiment_{it}$	$\overline{Sentiment}_{it-1,t-2}$	$MktRf_t$	SMB_t	HML_t	UMD_t	$FocalNews_{it}$	$PeerNews_{C_i t}$
Coeff	0.57	-0.01	0.0038	0.0028	0.19	0.31	-0.10	-0.32	0.0002	0.0002
(Robust SE)	(0.11)	(0.02)	(0.0011)	(0.0013)	(0.13)	(0.11)	(0.15)	(0.15)	(0.0007)	(0.0005)
Bootstrap *p*-value	0.00	0.61	0.01	0.05	0.08	0.01	0.81	1.00	0.45	0.39

H. Co-mention of News

To control for the potential co-mention news effect, we replace peer news in models (1) and (2) with co-mention news. The co-mention news data are retrieved from Factiva, which counts the number of news articles mentioning focal firms and other companies in global news repositories. We use the total number of news with co-mention of both focal stock and partner stock in the same search cluster and compute a co-mention index, $CoNews_{C_it} = {\sum_{j\in C_i, j\neq i} MktCap_{jt} \times \ln(1 + News_{ijt})}\Big/{\sum_{j\in C_i, j\neq i} MktCap_{jt}}$. $MktCap_{jt}$ is market capitalization of *j* at time *t* and $News_{ijt}$ is total number of news that mention both companies *i* and *j* at time *t*. We also replace Google News with Factiva news covering the total number of news associated with focal stock in $FocalNews_i$. Updated baseline and extended models can be stated as follows

$$R_{it} = \beta_{0i} + \beta_{1i}R_{C_it} + \beta_{2i}R_{C_it-1} + \beta_{3i}R_{S\&Pt} + \beta_{4i}R_{S\&Pt-1} + \beta_{5i}FocalNews_{it} + \beta_{6i}CoNews_{C_it} + \varepsilon_{it}. \quad \text{(i)}$$

$$R_{it} = \beta_{0i} + \beta_{1i}R_{C_it} + \beta_{2i}R_{C_it-1} + \beta_{3i}MktRf_t + \beta_{4i}SMB_t + \beta_{5i}HML_t + \beta_{6i}UMD_t + \beta_{7i}FocalNews_{it} + \beta_{8i}CoNews_{C_it} + \varepsilon_{it}, \quad \text{(ii)}$$

As shown in Tables G.1 and G.2, the main results do not change with inclusion of co-mention news. We still detect significant and positive comovement. The coefficient of cluster return in (i) is 0.55 and that in (ii) is 0.44. Focal news in (i) is positive and significant suggesting that the current return of focal firms is positively associated with global focal news volume. However, with controls of Fama-French four factors in (ii), the coefficient of focal news is insignificant. Co-mention news is also found to be insignificant in both (i) and (ii)

Table H.1. Comovement Regression with Co-mention News Volume: Baseline Model

	R_{C_it}	R_{C_it-1}	$R_{S\&Pt}$	$R_{S\&Pt-1}$	$FocalNews_{it}$	$CoNews_{C_it}$
Mean	0.55	0.02	0.64	-0.01	0.0011	-0.0007

(Robust SE)	(0.02)	(0.02)	(0.04)	(0.04)	(0.0005)	(0.0007)
Bootstrap *p*-value	0.00	0.21	0.00	0.43	0.03	0.17

Table H.2. Comovement Regression with Co-mention News Volume: Extended Model

	$R_{C_i t}$	$R_{C_i t-1}$	$MktRf_t$	SMB_t	HML_t	UMD_t	$FocalNews_{it}$	$CoNews_{C_i t}$
Mean	0.44	0.01	0.44	0.34	-0.18	-0.42	0.0007	-0.0007
(Robust SE)	(0.02)	(0.01)	(0.04)	(0.06)	(0.08)	(0.07)	(0.0005)	(0.0007)
Bootstrap *p*-value	0.00	0.08	0.00	0.00	0.04	0.00	0.15	0.16

I. Analysis with Control of Commonsensical Elements

Search clusters identified in our study may consist of commonsensical elements (e.g. competitor relationship and fundamental similarity). The comovement pattern may be driven by those elements. To address this concern, we re-run our main models excluding peers in our analysis of comovement. We also include placebo cluster returns to account for potential influence due to commonsensical elements.

Peers are identified by Yahoo! Finance based on the business nature of individual firms. In stock summary page, apart from "also-viewed" stocks, Yahoo! also shows a list of "comparison" stocks. We extract the "comparison" list and identify groups of peers with transitive relationship. For example, if B is a peer of A and C is a peer of B, we consider A, B and C to be peers though C does not appear in the "comparison" list of A. If a search cluster is formed by pure peer relationship, it is removed from the analysis. With removal of those search clusters, there are 291 stocks that form 63 clusters. Furthermore, in the analysis of comovement, we exclude peers of focal stocks in the calculation of R_{C_it} and R_{C_it-1}.

To account for fundamental similarity, we construct placebo cluster that consists of matching stocks that are similar in size, value, and industry as any of the cluster members and compute placebo cluster return R_{PC_it}. The original cluster members are excluded from the placebo cluster. We include the variable in our main models as shown in (i) and (ii) below.

$$R_{it} = \beta_{0i} + \beta_{1i}R_{C_it} + \beta_{2i}R_{C_it-1} + \beta_{3i}R_{PC_it} + \beta_{4i}R_{S\&Pt} + \beta_{5i}R_{S\&Pt-1} + \beta_{6i}FocalNews_{it} + \beta_{7i}CoNews_{C_it} + \varepsilon_{it}. \tag{i}$$

$$R_{it} = \beta_{0i} + \beta_{1i}R_{C_it} + \beta_{2i}R_{C_it-1} + \beta_{3i}R_{PC_it} + \beta_{4i}MktRf_t + \beta_{5i}SMB_t + \beta_{6i}HML_t + \beta_{7i}UMD_t + \beta_{8i}FocalNews_{it} + \beta_{9i}CoNews_{C_it} + \varepsilon_{it}, \tag{ii}$$

As shown in Tables H.1 and H.2, we still find positive and significant comovement with control of commonsensical elements.

Table I.1. Comovement Regression without Peers: Baseline Model

	R_{C_it}	R_{C_it-1}	R_{PC_it}	$R_{S\&Pt}$	$R_{S\&Pt-1}$	$FocalNews_{it}$	$CoNews_{C_it}$
Mean	0.54	0.02	0.50	0.06	-0.01	0.0008	-0.0011
(Robust SE)	(0.03)	(0.03)	(0.08)	(0.10)	(0.05)	(0.00)	(0.00)
Bootstrap *p*-value	0.00	0.08	0.00	0.20	0.30	0.02	0.15

Table I.2. Comovement Regression without Peers: Extended Model

	R_{C_it}	R_{C_it-1}	R_{PC_it}	$MktRf$	SMB_t	HML_t	UMD_t	$FocalNews_{it}$	$CoNews_{C_it}$
Mean	0.50	0.01	0.30	0.15	0.23	-0.16	-0.32	0.0005	-0.0009
(Robust SE)	(0.18)	(0.09)	(0.63)	(0.68)	(0.85)	(0.86)	(0.82)	(0.00)	(0.01)
Bootstrap *p*-value	0.00	0.03	0.00	0.13	0.00	0.01	0.00	0.09	0.27

J. Arbitrage Cost and Comovement

Prior studies suggest that return of a stock may be driven by arbitrage costs. In order to investigate whether comovement pattern is driven by arbitrage costs, we conduct further robustness check. We follow Kumar and Lee (2006) and Wurgler and Zhuravskaya (2002) and use variance of the residuals from a CAPM regression as a proxy for arbitrage cost. We use monthly stock returns 60 months before the time period of main analysis. We only consider stocks that experience cluster addition and analyze the effect of change in comovement. We divide all stocks into terciles and analyze comovement pattern in each tercile. As shown in Tables I.1 and I.2, the change in comovement in T3 is the highest.

Table J.1: Loading of Search Cluster Return in Baseline Model

	T1 (Lowest)	T2	T3 (Highest)
Coefficient	0.24	0.20	0.40
(Robust SE)	(0.07)	(0.09)	(0.11)
Bootstrap *p*-value	0.00	0.02	0.00

Table J.2: Loading of Search Cluster Return in Extended Model

	T1 (Lowest)	T2	T3 (Highest)
Coefficient	0.17	0.12	0.37
(Robust SE)	(0.06)	(0.10)	(0.12)
Bootstrap *p*-value	0.00	0.12	0.01

K. Pooled Regression Results

We repeat our main model and robustness tests (i) to (vi) using OLS regression with firm fixed effect and two dimensional cluster at firm and time levels. According to Petersen (2009), two-dimensional clustering produces better results for large panel data than OLS regression. The results are shown in Table J.1. Using pooled regression and with various controls, we still find positive and significant comovement among search clusters.

$$R_{it} = \beta_{0i} + \beta_{1i}R_{C_it} + \beta_{2i}R_{C_it-1} + \beta_{3i}R_{S\&Pt} + \beta_{4i}R_{S\&Pt-1} + \beta_{5i}FocalNews_{it} + \beta_{6i}PeerNews_{C_it} + \varepsilon_{it}. \quad \text{(i)}$$

$$\beta_{0i} + \beta_{1i}R_{C_it} + \beta_{2i}R_{C_it-1} + \beta_{3i}MktRf_t + \beta_{4i}SMB_t + \beta_{5i}HML_t + \beta_{6i}UMD_t + \beta_{7i}FocalNews_{it} + \beta_{8i}PeerNews_{C_it} + \varepsilon_{it}, \quad \text{(ii)}$$

$$R_{it} = \beta_{0i} + \beta_{1i}R_{C_it} + \beta_{2i}R_{C_it-1} + \beta_{3i}R_{S\&Pt} + \beta_{4i}R_{\&Pt-1} + \beta_{5i}FocalNews_{it} + \beta_{6}CoNews_{C_it} + \varepsilon_{it}. \quad \text{(iii)}$$

$$R_{it} = \beta_{0i} + \beta_{1i}R_{C_it} + \beta_{2i}R_{C_it-1} + \beta_{3i}MktRf_t + \beta_{4i}SMB_t + \beta_{5i}HML_t + \beta_{6i}UMD_t + \beta_{7i}FocalNews_{it} + \beta_{8i}CoNews_{C_it} + \varepsilon_{it}, \quad \text{(iv)}$$

$$R_{it} = \beta_{0i} + \beta_{1i}R_{C_it} + \beta_{2i}R_{C_it-1} + \beta_{3i}R_{PC_it} + \beta_{4i}R_{S\&Pt} + \beta_{5i}R_{S\&Pt-1} + \beta_{6i}FocalNews_{it} + \beta_{7i}CoNews_{C_it} + \varepsilon_{it}. \quad \text{(v)}$$

$$R_{it} = \beta_{0i} + \beta_{1i}R_{C_it} + \beta_{2i}R_{C_it-1} + \beta_{3i}R_{PC_it} + \beta_{4i}MktRf_t + \beta_{5i}SMB_t + \beta_{6i}HML_t + \beta_{7i}UMD_t + \beta_{8i}FocalNews_{it} + \beta_{9i}CoNews_{C_it} + \varepsilon_{it}, \quad \text{(vi)}$$

Table K.1. Pooled Regression Results

Model	(i)	(ii)	(iii)	(iv)	(v)	(vi)
R_{C_it}	0.50***	0.48***	0.50***	0.48***	0.45***	0.44***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
R_{C_it-1}	-0.01	0.00	-0.01	0.00	-0.01	-0.01
	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)
R_{PC_it}					0.52***	0.45***
					(0.06)	(0.06)
$R_{S\&Pt}$	0.69***		0.69***		0.14**	
	(0.05)		(0.05)		(0.06)	
$R_{S\&Pt-1}$	0.02		0.02		0.02	
	(0.03)		(0.03)		(0.02)	
$MktRf_t$		0.45***		0.46***		0.06
		(0.04)		(0.04)		(0.07)
SMB_t		0.27***		0.26***		0.16***
		(0.06)		(0.06)		(0.05)
HML_t		-0.14**		-0.14**		-0.11
		(0.07)		(0.07)		(0.07)
UMD_t		-0.34***		-0.33***		-0.24***
		(0.09)		(0.09)		(0.09)
$FocalNews_{it}$	-0.0000	-0.0000	0.0006*	0.0005	0.0006	0.0006
	(0.0005)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
$PeerNews_{C_it}$	-0.0002	0.0001				
	(0.0005)	(0.0004)				
$CoNews_{C_it}$			0.0000	-0.0002	-0.0001	-0.0002
			(0.0005)	(0.0005)	(0.0005)	(0.0005)
Constant	-0.0029**	-0.0036***	-0.0068***	-0.0057**	-0.0063***	-0.0056**
	(0.0014)	(0.0005)	(0.0022)	(0.0024)	(0.0022)	(0.0023)
	With Firm FE	With Firm FE	With Firm FE	With Firm FE	With Firm FE	With Firm FE
N	14,490	14,490	14,490	14,490	14,490	14,490
R^2	0.43	0.44	0.43	0.44	0.44	0.44

*** Significant at 1%,** Significant at 5%, * Significant at 10%

L. Dyadic Analysis

We conduct a dyadic analysis to explicitly determine the association between co-searching and price comovement across all possible stock pairs. We consider all possible pairs among the Russell 3000 stocks and determine whether or not they belong to the same search cluster in the month of September 2011. We test whether or not the co-searching of a stock pair can be associated with the weekly price correlation between the stocks in the pair in the subsequent two months. We run the following pooled regression to test this association

$$Corr_{i,j,t} =$$

$$\beta_0 + \beta_1 SameCluster_{i,j} + \beta_2 DiffMktCap_{i,j,t-1} + \beta_3 DiffP2B_{i,j,t-1} + \beta_4 SameFF10_{i,j,t-1} +$$

$$\beta_5 VolChange_{i,j,t-1} + \beta_6 Corr_{i,j,t-1} + \beta_i \delta_i + +\beta_j \delta_j + \varepsilon_{it}$$

where $Corr_{i,j,t}$: Person's correlation of daily returns between stocks *i* and *j* in week *t*

$SameCluster_{i,j}$: Indicator variable which is 1 if stocks *i* and *j* are in the same search cluster in September 2011 and 0 otherwise

$DiffMktCap_{i,j,t-1}$: Absolute difference in average market capitalization between stocks *i* and *j* in week *t-1*

$DiffP2B_{i,j,t-1}$: Absolute difference in average price to book ratio between stocks *i* and *j* in week *t-1*

$SameFF10_{i,j,t-1}$: Indicator variable which is 1 if stocks *i* and *j* belong to the same FF10 in week t-1 and 0 otherwise

$VolChange_{i,j,t-1}$: Absolute difference in percentage change in trading volume from week *t*-2 to week *t*-1 between stocks *i* and *j*

$Corr_{i,j,t-1}$: Person's correlation of daily returns between stocks *i* and *j* in week *t-1*

δ_i & δ_j: dummy variables for stocks i and j respectively

We estimate the above equation using OLS. In order to account for cross-sectional and time series correlation we also apply two dimensional clustering (Peterson 2009). The results are shown in Table K.1.

The coefficient for $SameCluster_{i,j}$, is positive and significant. Our results confirm that the co-searching of stocks is associated with the subsequent price comovement.

Table L.1. Regression Result of Dyadic Analysis

Variable	Model 1	Model 2
$SameCluster_{i,j}$	0.067***	0.073***
	(0.009)	(0.010)
$Corr_{i,j,t-1}$		-0.062
		(0.038)
$VolChange_{i,j,t-1}$	-0.002	-0.005
	(0.005)	(0.006)
$DiffMktCap_{i,j,t-1}$	-0.000**	-0.000**
	(0.000)	(0.000)
$DiffP2B_{i,j,t-1}$	-0.002**	-0.002**
	(0.001)	(0.001)
$SameFF10_{i,j,t-1}$	0.014**	0.015**
	(0.004)	(0.005)
Constant	0.950**	0.975***
	(0.251)	(0.252)
	With Firm Fixed Effect	With Firm Fixed Effect
N	9,058,408	9,058,060
R^2	0.14	0.15

*** Significant at 1%,** Significant at 5%, * Significant at 10%

M. Comovement Analysis with different Top Ranking Lists of Co-viewing Stocks

Table M.1. Comovement Regression Estimates: Baseline Model

Panel A. Top 5 Co-Searching Partners

	R_{C_it}	R_{C_it-1}	$R_{S\&Pt}$	$R_{S\&Pt-1}$	$FocalNews_{it}$	$PeerNews_{C_it}$
Mean	0.55	0.03	0.62	-0.04	-0.0003	-0.0001
(Robust SE)	(0.02)	(0.02)	(0.03)	(0.03)	(0.0005)	(0.0004)
Bootstrap *p*-value	0.00	0.09	0.00	0.12	0.3450	0.4310

Panel B. Top 4 Co-Searching Partners

	R_{C_it}	R_{C_it-1}	$R_{S\&Pt}$	$R_{S\&Pt-1}$	$FocalNews_{it}$	$PeerNews_{C_it}$
Mean	0.08	-0.05	1.22	0.04	-0.0003	-0.0004
(Robust SE)	(0.03)	(0.03)	(0.05)	(0.05)	(0.0006)	(0.0005)
Bootstrap *p*-value	0.00	0.07	0.00	0.19	0.3890	0.2720

Panel C. Top 3 Co-Searching Partners

	R_{C_it}	R_{C_it-1}	$R_{S\&Pt}$	$R_{S\&Pt-1}$	$FocalNews_{it}$	$PeerNews_{C_it}$
Mean	0.54	0.02	0.61	-0.04	-0.0005	0.0000
(Robust SE)	(0.03)	(0.03)	(0.04)	(0.04)	(0.0007)	(0.0005)
Bootstrap *p*-value	0.00	0.21	0.00	0.15	0.3460	0.3670

Table M.2. Comovement Regression Estimates: Extended Model

Panel A. Top 5 Co-Searching Partners

	R_{C_it}	R_{C_it-1}	$MktRf_t$	SMB_t	HML_t	UMD_t	$FocalNews_{it}$	$PeerNews_{C_it}$
Mean	0.47	0.02	0.47	0.27	-0.08	-0.31	-0.0003	0.0002
(Robust SE)	(0.02)	(0.01)	(0.04)	(0.06)	(0.08)	(0.07)	(0.0005)	(0.0004)
Bootstrap *p*-value	0.00	0.02	0.00	0.00	0.21	0.00	0.3080	0.3400

Panel B. Top 4 Co-Searching Partners

	R_{C_it}	R_{C_it-1}	$MktRf_t$	SMB_t	HML_t	UMD_t	$FocalNews_{it}$	$PeerNews_{C_it}$
Mean	0.07	-0.01	0.87	0.39	-0.11	-0.47	-0.0002	0.0001
(Robust SE)	(0.03)	(0.02)	(0.07)	(0.10)	(0.12)	(0.12)	(0.0006)	(0.0004)
Bootstrap *p*-value	0.05	0.24	0.00	0.00	0.08	0.00	0.4740	0.4850

Panel C. Top 3 Co-Searching Partners

	R_{C_it}	R_{C_it-1}	$MktRf_t$	SMB_t	HML_t	UMD_t	$FocalNews_{it}$	$PeerNews_{C_it}$
Mean	0.44	0.00	0.48	0.16	-0.12	-0.38	-0.0004	0.0002
(Robust SE)	(0.03)	(0.02)	(0.06)	(0.08)	(0.12)	(0.11)	(0.0007)	(0.0005)
Bootstrap *p*-value	0.00	0.35	0.00	0.03	0.23	0.00	0.3610	0.5230

N. Dynamic Comovement with Placebo Control

We want to evaluate the change in the comovement of the stock as it is added or deleted from a search cluster and compare it with the change in the comovement of a placebo stock or a control stock. In order to do so we follow the approach used by Boyer (2011) and compared the difference in the comovement coefficients of focal and placebo stocks, accompanying an addition or deletion event. More specifically, we estimate models 3A, 4A, 3B, and 4B for each focal stock *i* that experiences addition/removal and its corresponding placebo stock that matches stock *i* in the same FF10 industry and closet size and value. We determine the difference in comovement of focal and placebo stocks based on the difference in the coefficient changes for these stocks before and after the event i.e. $\left(\beta_{i=focal}^{t_4} - \beta_{i=focal}^{t_2}\right) - \left(\beta_{i=placebo}^{t_4} - \beta_{i=placebo}^{t_2}\right)$.

Results for addition and deletion are showing in Table M1. As shown in Table M1, after addition, the focal stock has a more positive change in the coefficient as compared to the placebo stock and this difference is significant i.e. $\Delta\Delta R_{C_i t}$ is positive and significant. Similarly, after deletion, the focal stock has a more negative change in the coefficient and this difference is significant i.e. $\Delta\Delta R_{C_i t}$ is negative and significant in Table M1. The results are consistent to our main analysis that focal stocks co-move more with the peers when they are in the same search clusters and their comovement is much higher with the cluster stocks as compared to the comovement of placebo stocks.

Table N.1. Dynamic Comovement Analysis Results

Panel A: Summary Statistics of Estimated Betas (Additions)

N=95	Model	$\Delta\Delta R_{C_i t}$	$\Delta\Delta R_{C_i t-1}$	$\Delta\Delta R_{S\&Pt}$	$\Delta\Delta R_{S\&Pt-1}$	$\Delta\Delta MktRf$	$\Delta\Delta SMB$	$\Delta\Delta HML$	$\Delta\Delta UMD$	$\Delta\Delta FocalNews_i$	$\Delta\Delta PeerNews$
Mean (SE)	A	0.54 (0.14)	-0.13 (0.15)	-2.77 (0.56)	-0.29 (0.55)					0.0022 (0.0020)	-0.0007 (0.0014)
	B	0.34 (0.03)	-0.01 (0.03)			-0.55 (0.07)	-0.31 (0.07)	0.10 (0.11)	-0.16 (0.10)	0.0008 (0.0003)	0.0011 (0.0004)
p-value	A	0.00	0.22	0.00	0.22					0.28	0.38
	B	0.00	0.48			0.00	0.06	0.23	0.33	0.33	0.28

Panel B: Summary Statistics of Estimated Betas (Deletions)

N=205	Model	$\Delta\Delta R_{C_i t}$	$\Delta\Delta R_{C_i t-1}$	$\Delta\Delta R_{S\&Pt}$	$\Delta\Delta R_{S\&Pt-1}$	$\Delta\Delta MktRf$	$\Delta\Delta SMB$	$\Delta\Delta HML$	$\Delta\Delta UMD$	$\Delta\Delta FocalNews_i$	$\Delta\Delta PeerNews$
Mean (SE)	A	-0.08 (0.02)	-0.03 (0.02)	0.25 (0.04)	0.1624 (0.0591)					-0.0005 (0.0007)	0.0009 (0.0006)
	B	-0.07 (0.02)	0.00 (0.01)			0.29 (0.05)	-0.14 (0.03)	0.13 (0.07)	0.11 (0.04)	-0.0001 (0.0007)	0.0006 (0.0006)
p-value	A	0.00	0.31	0.00	0.09					0.23	0.05
	B	0.04	0.42			0.00	0.11	0.18	0.05	0.49	0.13

O. Prediction Model with Control for the Effect of Sentiments

We repeat the prediction analysis in section 7 with additional control for the effect of sentiments. We determine the sentiment scores of positive and negative articles associated with each stock on popular sites such as Seeking Alpha, InvestorWired.com, WallStreet.org, and Forbes from the site thestocksonar.com. We separately control for the positive sentiment $ln(Pos_{i,t-6:t-1})$ and negative sentiment $ln(Neg_{i,t-6:t-1})$ associated with the focal stock. Results are shown in Table N.1 and are qualitatively similar to our main prediction analysis.

Table O.1. Prediction of 1 Month Future Returns

Panel A. Regression Results

	M1	M2	M3	M4
$R_{C_i,t-6:t-1}$	0.049**	0.051**	0.029***	0.029**
	(0.021)	(0.021)	(0.011)	(0.011)
$R_{i,t-6:t-1}$	0.032**	0.032**	0.017**	0.018**
	(0.013)	(0.013)	(0.008)	(0.008)
$FF49_{i,t-6:t-1}$	0.000	0.004	-0.004	-0.004
	(0.016)	(0.015)	(0.012)	(0.012)
$ln(\overline{MktCap}_{i,t-6:t-1})$	-0.003	-0.002	-0.002	-0.001
	(0.004)	(0.004)	(0.003)	(0.003)
$ln(\overline{P2B}_{i,t-6:t-1})$	-0.005	-0.005	-0.003	-0.003
	(0.010)	(0.010)	(0.003)	(0.003)
$\overline{Leverage}_{i,t-6:t-1}$	0.053**	0.047**	0.053***	0.053***
	(0.022)	(0.022)	(0.010)	(0.010)
$R_{i,t-36:t-1}$	0.006	0.006	-0.001	-0.001
	(0.004)	(0.004)	(0.003)	(0.003)
$ln(\overline{News}_{i,t-6:t-1})$	0.007**	0.007***	0.008***	0.009***
	(0.003)	(0.002)	(0.002)	(0.002)
$ln(SVI_{i,t-6:t-1})$	-0.003		-0.002	
	(0.002)		(0.001)	
$ln(ASVI_{i,t-6:t-1})$		-0.021**		-0.019***
		(0.009)		(0.005)
$ln(Pos_{i,t-6:t-1})$	0.002	-0.001	0.009***	0.007**
	(0.005)	(0.005)	(0.003)	(0.003)
$ln(Neg_{i,t-6:t-1})$	-0.002	-0.001	-0.007***	-0.008***
	(0.003)	(0.003)	(0.002)	(0.002)
$ln(Vol_{i,t-6:t-1})$	0.002	0.003	0.002	0.001
	(0.003)	(0.003)	(0.002)	(0.002)
$ln(Turnover_{i,t-6:t-1})$	-0.004	-0.003	-0.003	-0.003
	(0.005)	(0.005)	(0.004)	(0.004)
$Volatility_{i,t-6:t-1}$	0.061	0.180	0.383	0.566**
	(0.360)	(0.384)	(0.250)	(0.256)
Constant	-0.097**	-0.093**	-0.111***	-0.096**

	(0.043)	(0.045)	(0.042)	(0.043)
N	7,876	7,876	7,876	7,876
R^2	0.2579	0.2592	0.0311	0.0329
Model Type	FM	FM	TC	TC

***Significant at 1%, ** Significant at 5%, * Significant at 10%
FM: Fama-MacBeth Regression with Newey-West correction of a lag of 4
TC: OLS Regression with two dimensional clustering at firm and time level

Panel B. Cluster without Commonsensical Peers Regression Results

	M1	M2	M3	M4
$R_{C_i,t-6:t-1}$	0.047**	0.049**	0.029***	0.028**
	(0.020)	(0.020)	(0.011)	(0.011)
$R_{i,t-6:t-1}$	0.032**	0.032**	0.017**	0.018**
	(0.013)	(0.013)	(0.008)	(0.008)
$FF49_{i,t-6:t-1}$	0.002	0.007	-0.007	-0.007
	(0.016)	(0.015)	(0.013)	(0.013)
$ln(\overline{MktCap}_{i,t-6:t-1})$	-0.004	-0.003	-0.001	-0.001
	(0.004)	(0.004)	(0.003)	(0.003)
$ln(\overline{P2B}_{i,t-6:t-1})$	-0.005	-0.005	-0.003	-0.004
	(0.011)	(0.011)	(0.004)	(0.004)
$\overline{Leverage}_{i,t-6:t-1}$	0.062***	0.056**	0.065***	0.064***
	(0.021)	(0.023)	(0.011)	(0.011)
$R_{i,t-36:t-1}$	0.006	0.007*	-0.001	-0.001
	(0.004)	(0.004)	(0.003)	(0.003)
$ln(\overline{News}_{i,t-6:t-1})$	0.008***	0.008***	0.008***	0.009***
	(0.003)	(0.002)	(0.002)	(0.002)
$ln(SVI_{i,t-6:t-1})$	-0.003		-0.003*	
	(0.003)		(0.001)	
$ln(ASVI_{i,t-6:t-1})$		-0.025**		-0.020***
		(0.009)		(0.006)
$ln(Pos_{i,t-6:t-1})$	0.003	0.000	0.010***	0.008***
	(0.005)	(0.005)	(0.003)	(0.003)
$ln(Neg_{i,t-6:t-1})$	-0.004	-0.003	-0.009***	-0.010***
	(0.004)	(0.004)	(0.003)	(0.003)
$ln(Vol_{i,t-6:t-1})$	0.003	0.004	0.002	0.001
	(0.003)	(0.003)	(0.003)	(0.003)
$ln(Turnover_{i,t-6:t-1})$	-0.005	-0.004	-0.001	-0.001
	(0.005)	(0.005)	(0.004)	(0.004)
$Volatility_{i,t-6:t-1}$	0.057	0.197	0.375	0.561**
	(0.353)	(0.391)	(0.260)	(0.267)
Constant	-0.112**	-0.117**	-0.111**	-0.100**
	(0.053)	(0.056)	(0.050)	(0.050)
N	6,844	6,844	6,844	6,844
R^2	0.2744	0.2749	0.0350	0.0367
Model Type	FM	FM	TC	TC

Online Appendix References

Antweiler, W., Frank, M.Z. (2004) Is all that talk just noise? The information content of Internet stock message boards. *The Journal of Finance* 59(3): 1259-1294.

Boyer, B.H. (2011) Style-related comovement: Fundamentals or labels? *The Journal of Finance* 66(1): 307-332.

Dehning, B., Richardson, V.J., Zmud, R.W. (2007) The financial performance effects of IT-based supply chain management systems in manufacturing firms. *Journal of Operations Management* 25(4): 806-824.

Fellner, G., Maciejovsky, B. (2007) Risk attitude and market behavior: Evidence from experimental asset markets. *Journal of Economic Psychology* 28(3): 338-350.

Greene, W.H. (2003) *Econometric analysis* (Prentice Hall, Upper Saddle River, NJ).

Kumar, A. (2009) Who gambles in the stock market? *The Journal of Finance* 64(4): 1889-1933.

Kumar, A., Lee, C.M.C. (2006) Retail investor sentiment and return comovements. *The Journal of Finance* 61(5): 2451-2486.

Kumar, A., Page, J.K., Spalt, O.G. (2011) Religious beliefs, gambling attitudes, and financial market outcomes. *Journal of Financial Economics* 102(3): 671-708.

Menzly, L., Ozbas, O. (2010) Market segmentation and cross-predictability of returns. *The Journal of Finance* 65(4): 1555-1580.

Patton, A., Politis, D., White, H. (2009) Correction to “automatic block-length selection for the dependent bootstrap” by D. Politis and H. White. *Econometric Reviews* 28(4): 372-375.

Petersen, M.A. (2009) Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies* 22(1): 435-480.

Politis, D.N., White, H. (2004) Automatic block-length selection for the dependent bootstrap. *Econometric Reviews* 23(1): 53-70.

Sabherwal, S., Sarkar, S.K., Zhang, Y. (2011) Do Internet stock message boards influence trading? Evidence from heavily discussed stocks with no fundamental news. *Journal of Business Finance & Accounting* 38(9-10): 1209-1237.

Terwiesch, C., Ren, Z.J., Ho, T.H., Cohen, M.A. (2005) An empirical analysis of forecast sharing in the semiconductor equipment supply chain. *Management Science* 51(2): 208-220.

Wurgler, J., Zhuravskaya, E. (2002) Does arbitrage flatten demand curves for stocks? *The Journal of Business* 75(4): 583-608.