The Dissertation Committee for Chung Man Alvin Leung Certifies that this is the approved version of the following dissertation:

Attention, Search, and Information Diffusion: Study of Stock Network Dynamics and Returns

Committee:

Prabhudev Konana, Supervisor

Ashish Agarwal, Co-Supervisor

Andrew B. Whinston

Anitesh Barua

Alok Kumar

Attention, Search, and Information Diffusion: Study of Stock Network Dynamics and Returns

by

Chung Man Alvin Leung, BBA; BE; MPHIL; MSIROM

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2014

Dedication

To my late great grandmother for her inspiration

And

To my family for their continuous support and encouragement

Acknowledgements

It has been a long journal for my PhD study. I thank everyone who has offered me help and provided accompanied me to overcome so many hurdles. First and foremost, I would like to express my profound gratitude to my supervisor, Prabhudev Konana, for his patience and guidance in the past five years. He is always my source of inspiration, my guide and mentor. I am also indebted to my co-supervisor Ashish Agarwal for his advice and encouragement. His knowledge and experience have contributed a lot to my PhD study. I would also like to express my heartfelt gratitude to my dissertation committee members, Andrew Whinston, Anitesh Barua and Alok Kumar for their advice and feedback on my dissertation. I am indebted to Sirkka Javenpaa and other faculty members who have taught me in the past five years. They have extended my knowledge in academic research and stimulate my intellectual curiosity.

It is an honor to be a PhD student of McCombs School of Business. I wholeheartedly thank the School and IROM department for their financial support and resources to help me accomplish my dissertation. I am also grateful to Supply Chain Management Center for Excellence who provided me with additional financial support to attend various conferences to present my research.

Last but not least, I would like to express my gratitude to the graduate coordinators, Caroline Walls and Hillary Patterson, who helped me a lot throughout the PhD life, my PhD friends who have given me a lot of encouragement and spiritual support, and my parents who have sacrificed a lot in the past five years.

Attention, Search, and Information Diffusion: Study of Stock Network Dynamics and Returns

Chung Man Alvin Leung, PhD

The University of Texas at Austin, 2014

Supervisors: Prabhudev Konana and Ashish Agarwal

There is growing literature on search behavior and using search for prediction of market share or macroeconomic indicators. This research explores investors' stock search behaviors and investigates whether there are patterns in stock returns using those for return prediction. Stock search behaviors may reveal common interest among investors. In the first study, we use graph theory to find investment habitats (or search clusters) formed by users who search common set of stocks frequently. We study stock returns of stocks within the clusters and across the clusters to provide theoretical arguments that drive returns among search clusters. In the second study, we analyze return comovement and cross-predictability among economically related stocks searched frequently by investors. As search requires a considerable amount of cognitive resources of investors, they only search a few stocks and pay high attention to them. According to attention theory, the speed of information diffusion is associated with the level of attention. Quick information diffusion allows investors to receive relevant information immediately and take instantaneous trading action. This immediate action may lead to correlated return comovement. Slow information diffusion creates latency between the occurrence of an event and the action of investors. The slower response may lead to cross-predictability. Making use of the discrepancy in information diffusion, we implement a trading strategy

to establish arbitrage opportunities among stocks due to difference in user attention. This research enriches the growing IS literature on information search by (1) identifying new investment habitats based on user search behaviors, (2) showing that varying degrees of co-attention and economic linkages may lead to different speed of information diffusion (3) developing a stock forecasting model based on real-time co-attention intensity of a group economically linked stocks and (4) embarking a new research area on search attention in stock market. The methods in handling complex search data may also contribute to big data research.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Motivation

"Never pay the slightest attention to what a company president ever says about his stock." – Bernard Baruch

There is much interest to build a strong foundation around search behavior, frequency, and trends (e.g., Google search and trends) and market prediction. Wu and Brynjolfsson (2009) use search engine data to predict future housing market sales and prices. Choi and Varian (2012) show how to use Google Trends to forecast near-term values of economic indicators. More recently, Pries et al. (Preis et al. 2013) use Google Trends data to predict market movement. These studies that rely on search volume and trends differ from those using text sentiment analysis of daily news articles and participation in message board postings to predict market movement (Antweiler and Frank 2004, Lam 2004, Tetlock 2007, Tsai and Hsiao 2010).

Motivated by the growing popularity in search, we study the stock price search behaviors of online investors and investigate the implications of such behaviors on information diffusion and returns prediction. As active search requires cognitive resources, the intensity of search behaviors may reflect the level of attentiveness of investors. Attention theory in finance attributes many abnormalities in stock returns to the lack of user attention. However, such theory lacks empirical support. This research using search behavior may provide empirical support to the attention theory. If search behaviors can foreshadow stock movement, new insights and theories on information search can be developed. Furthermore, in prior studies on stock prediction, most models are built on post-hoc trading data. In this research, we build a prediction model based on pre-transaction data. This study is important to IS research. We may demonstrate that search behaviors of online users are not random. They have important implications to return comovement and stock prediction. Furthermore, the development of real-time prediction model may enrich existing design science research on stock prediction.

This dissertation is interdisciplinary in nature and at the intersection of search behavior, psychology of attention, and finance. This research is consistent with numerous other streams of recent research in IS investigating search behavior at the intersection of IS and marketing, and IS and economics.

1.2 OVERVIEW

The dissertation consists of two studies in following chapters. Study 1 is to identify investment habitats formed by investors' search behaviors. Study 2 investigates the speed of information diffusion among stock with disparate level of co-search intensity. Based on the discrepancy in information diffusion, I develop a prediction model based on investors' search behaviors.

In the first study, I analyze the common interest of stocks among investors on Yahoo! Finance and identify different clusters of stocks based on their search behaviors for various stocks. Each cluster reveals special interests of potential investors. I investigate what drives the formation of each cluster and their implications on stock movement. As investors' search behaviors change over time, I also analyze the dynamic changes in the clusters and the impact on stock movement.

In the second study, I analyze information diffusion across publicly listed firms with different degrees of economic linkages and co-attention intensity. In finance, many theories have been built around attention, but yet empirically demonstrated. That is, finance literature uses attention theory to explain stock movements, but they do not demonstrate the attention or inattention. For example, prior studies (e.g. Cohen and Frazzini 2008, Menzly and Ozbas 2010) posit that lead-lag effects exist among suppliers and buyers due to inattention. Comovement theories also suggest that stock returns comove among stocks in the same categories (e.g. Big Cap, Small Cap, Growth Stocks, Value Stocks) (Barberis et al. 2005). This research may provide empirical support to the inattention and comovement theories based on search attention for different stocks with economic linkages. Capitalizing on the slow information diffusion among low attention supply-chain related stocks, we implement a trading strategy. A positive portfolio return is obtained by buying stocks of firms whose associated peers had significant positive returns in the previous week and selling stocks of firms whose associated peers had significant negative returns in the previous week. Due to slow information diffusion, the positive/negative information of associated firms will reach target firms with lagged time. This creates an arbitrage opportunity to seek higher returns.

There are several contributions of this research. First, we can identify investment habitats of interest to investors. The identification method of investment habitats is different from traditional finance research that uses post-hoc trading data to figure out areas of interest to investors. For example, in the study of style investing[1], researchers analyze investment portfolios (e.g. small, big, growth, value) based on the transaction

[1] Style investing is an investment approach of investors who invest in stocks with certain style (e.g., Big - cap, small-cap, growth stocks, value-tocks, international stocks, emerging market stocks). They may change their investment style based on the historical performance of stocks in the same category.

data of some brokerage firms (Froot and Teo 2008, Kumar 2009a). Different from prior finance studies, the data used in this research are based on search behaviors of millions of investors. The results seem to be more robust based on larger sample size. Furthermore, trading is limited to investors with sufficient financial capabilities. However, searching is not restricted by any financial constraints. Therefore, the search data can reflect the interest of a broader range of investors. Due to larger sample data, the investment habitats identified in this research may better represent preferences of investors than the broad categories defined in prior research on style investing.

Additionally, this research may demonstrate that search can be used as a proxy for attention. In finance research, attention is coarsely measured by news coverage, extreme stock returns and trading volume. It is doubtful whether reaction in the market is a precise measure of the most recent user attention. Also, some non-popular stocks are not reported in news as frequently as other mainstream stocks. Therefore, news coverage seems to be biased to popular stocks. Furthermore, search provides real-time data to measure attention of the majority of investors unlike the use of post-hoc data in the previous results.

In addition, this research may contribute to the understanding of the driving forces of information diffusion. Both the strength of economic linkages and the intensity of co-attention may play important roles in information diffusion, which lead to return comovement or cross-predictability. Furthermore, based on the conditions leading to slow information diffusion, it is possible to develop stock return prediction models. In prior studies, prediction models are based on aggregated market/industry data. In this research, we show that it may be possible to use more granular firm level data to make a prediction. Making use of co-attention intensity of economically linked firms, it is possible to develop a trading strategy to seek arbitrage benefits in real-time. Finally, as search is frequent in the Internet era, we may apply the same research model to analyze other commodity products with correlated demands. This research may embark a new area of study on information search.

Chapter 2: Network Analysis of Search Dynamics: The Case of Stock Habitats

2.1 Introduction

Online search has attracted significant attention among IS researchers because it has implications for economic outcomes such as prices and market efficiency (Bakos 1997, Brynjolfsson and Smith 2000, Granados et al. 2012, Kuruzovich et al. 2008, Weber and Zheng 2007). Researchers have used such aggregate search trends to predict house prices (Wu and Brynjolfsson 2009), near-term economic indicators (e.g., unemployment rate) (Choi and Varian 2012), stock market movement (Preis et al. 2013), and firm equity values (Luo et al. 2013). This research expands this body of knowledge where users conduct search on a set of products or assets (i.e., searches are correlated) to reveal underlying characteristics of that set. This study is important to IS by (1) identifying dynamic changes of user preferences over time from co-search networks (2) showing the linkages between search and future demand correlation, and (3) extending the growing literature on network economy and demonstrating the collective inference and economic influence of search based networks.

A few studies have recently focused on correlated search activities of multiple items (e.g., Kim et al. 2011). Due to the enormous amount of cognitive resources needed for search, users focus their attention on a limited set of products (Kahneman 1973, Li et al. 2013). At the same time, due to heterogeneity in preferences, different groups of users would focus on different sets of products/assets at a given time. As a result, correlated searches reveal user attention to a limited set of items (e.g., products or stocks) at a particular time window. These correlated search sets may change over time as users' attention evolves. As IT platforms capture consumer digital footprints (i.e., search history), we can now analyze aggregated search correlations of all users to extract a number of insights. If search correlations across items are associated with subsequent demand correlations across items, then search can reveal aggregated user preferences at any given time. Such user preferences information can be used to make market-level predictions (Kim et al. 2011). The insights from correlated searches allow appropriate strategic actions, including pricing, promotions, and inventory management. For example, knowledge of correlation in user preferences for products or brand can be used for co-promotions and to form partnerships.

Past studies have used transaction-level data to evaluate demand correlations across products. Researchers have used data mining methods such as market basket analyses to analyze such transaction-level data to understand correlations among different products and product categories (Manchanda et al. 1999, Mehta 2007, Niraj et al. 2008). Oestreicher-Singer and Sundararajan (2012a, 2012b) analyze the effect of IT platform on understanding the impact of demand correlations on the demand for *individual* products. However, these studies do not consider search correlations and the role of search in revealing demand correlations. Though we do not claim that search network is better than product network in demand prediction, the use of search network has several advantages. First, product network is based on transactional data which are available after transactions are complete. As search precedes transactions, the analysis of search network may have time advantages in identifying customer preferences at an earlier time period.

Insights derived from search networks can help companies develop appropriate marketing strategies to convert web visitors to customers. Second, transactional data captures the final decision of customers in their purchase process. They may not reveal the consideration sets of customers. When people search, they evaluate products whose product features may meet their expectation. Search data allow us to identify consideration sets of customers. Third, transactional data are proprietary to retailers. They may not be free to the public. In contrast, search data may be easier to access as search activities can be easily tracked by public IT platforms (e.g. search engine and forums).

While literature is evolving to understand correlated searches, the key is to recognize that correlated searches of products or assets (e.g., stocks) form a complex network of interconnectedness based on many different dimensions and contexts. For instance, for physical products, these dimensions could be related to price and product attributes. In the context of stocks, these dimensions may be related to firm size, industry, volatility, supply chain relationships, and other financial measures. Several recent studies have focused on user-based networks and their economic implications for users and products such as YouTube video clips (Susarla et al. 2012), blogs (Mayzlin and Yoganarasimhan 2012), news reports (Calin et al. 2012), photos (Zeng and Wei 2013), and loans (Lin et al. 2013). However, we can develop interesting insights for decision-makers by extracting information from a network of correlated searches. Sundararajan et al. (2013) argue that there is much interest in network economy research in studying how IT creates and reveals networks, recognizing the economic impacts of information flows through the network, and understanding network-based inferences and dynamics. Thus, combining correlated searches and network analysis can provide deeper insights on the underlying behavior of products or assets.

We extend this evolving literature by exploring the use of search to address the following research questions: (a) Can network analysis of correlated searchers reveal underlying behavior of products or assets from which we can infer users' preferences?, (b) How does changing user attention affect search network dynamics?, (c) Do the underlying behaviors of products or assets change with changing attention?, and (d) How does correlated search network help in predicting future demand?

While past literature has generally focused on physical products, very little has been studied to understand the search networks in the context of finance. Millions of investors flock to IT portals such as Yahoo! Finance to search stock prices. Collective searches on such platforms could potentially provide deeper insights into investor preferences and the underlying behaviors of the stocks (e.g., returns of the set of stocks searched). There is a growing body of literature in finance related to investment preferences, called investment habitats, where investors tend to pay attention to a small set of stocks based on many different dimensions such as size, industry, geography, etc. Finance literature shows that returns tend to comove within these habitats (Barberis et al. 2005). Since investors search many different stocks together, the collective correlated searches can reveal search clusters and their properties.[2] The search clusters can change over time as investors focus on different sets of stocks. However, it is unclear whether the

[2] We do not make any assumptions that searches result in actual trading or comovements. Investors search for stock information for monitoring, curiosity, or information gathering prior to investment. The focus of this research is to see if collective search data can reveal return outcomes rather than influencing the trading decisions. We do not observe or assume any trading decisions.

information-seeking behavior reflects actual investment habitats and the properties of habitats. Additionally, it is not obvious if changes in search clusters over time also reveal the changes in the return correlation of the associated stocks. Insights from such analysis can have significant implications for cross predictability of stocks within the same habitats or between habitats. For example, Wahal and Yavuz (2013) show that high comovement momentum portfolios have significantly higher future returns than low comovement momentum portfolios. Thus, we can explore the role of search in revealing investment habitats and return comovement. Further, stock data are publicly available, and as a result we can easily validate if search correlations among stocks reflect return correlations.

Using the online search data from Yahoo! Finance of Russell 3000 index stocks, we construct a correlated search network using graph theory, where the nodes represent stocks and edges represent co-search strength. Using the graph theoretic approach, we extract cliques and non-overlapping clusters and analyze the underlying behavior of stock clusters. We found 50 to 79 search clusters at different points in time representing 230 to 349 stocks. Surprisingly, most of the stocks in the Russell 3000 index do not belong to any search cluster. However, the stock returns within these search clusters are strongly correlated. We control for the effects of commonly used risk factors and news related to the stocks in the cluster. Even after accounting for these known determinants of return comovement, we find that the returns of stocks in the same search cluster are strongly correlated consistent with the predictions of the habitat-based framework for return comovement proposed in Barberis, Shleifer, and Wurgler (2005).

We find that the stocks within clusters change over time and the comovement patterns also change after controlling for all known stock characteristics (e.g., size, price-to-book ratio, industry, volatility, and supply-chain relationships). More specifically, when a stock enters (departs) a search cluster, the focal stock return comoves (detaches) with the cluster returns. Thus, changing search clusters can reveal changing investment habitats. This is an interesting result, since individual investors may not have any influence on stock prices, but collectively the searches may provide interesting insights regarding the investor preferences.

This research has important managerial implications. We show that search on IT platforms may be used to extract user preferences for products or assets with interesting underlying behavior. The methodology discussed here can help improve the design of recommendation systems or collectively intelligent systems on IT portals or e-commerce sites and can be used in different contexts such as selling books, movies, or electronic appliances. Based on the search preferences of online users, marketers can promote their products to potential customers more effectively. Additionally, our approach of forming search clusters and validating comovement can be used by the online platforms as a feedback system to evaluate and validate preferences based on search.

Our paper contributes to the existing research in several ways. Previous studies have focused primarily on evaluating customer preferences at the individual level (Atahan and Sarkar 2011, Moe and Fader 2004) and developing methods for recommendations based on aggregate preferences using transactional data (Ansari et al. 2000, Huang et al. 2007, Van Roy and Yan 2010). This paper demonstrates that correlated searches can be used to determine aggregate preferences over time on a real-time basis. Our proposed methodology may help in understanding dynamic preferences of consumers and complement existing research that extract insights from blogs (Aggarwal et al. 2012, Dewan and Ramaprasad 2012, Dhar and Chang 2009, Droge et al. 2010), reviews and ratings (Chevalier and Mayzlin 2006, Dellarocas et al. 2007, Duan et al. 2008a, Duan et al. 2008b, Godes and Mayzlin 2004, Gu et al. 2012, Zhu and Zhang 2010).

Finally, our study has important implications for finance research on attention and investment habitats. Previous studies (e.g., Da et al. 2011, Preis et al. 2013) use Google Trends search volume as a proxy for investor attention and analyze the impact of attention on individual stock returns. Our study extends their research and proposes a measure of investor attention simultaneously toward a group of stocks. Furthermore, prior investment habitat studies (e.g., Froot and Teo 2008, Graham and Kumar 2006, Green and Hwang 2009, Greenwood 2008, Huberman and Dorn 2009, Kumar et al. 2013, Pindyck and Rotemberg 1993, Pirinsky and Wang 2006, Vijh 1994) primarily focus on fundamental habitat. We demonstrate that investment habitats may change over time and that such change may not be fully captured by firm fundamentals (e.g. size, value, and industry). Furthermore, we find that search clusters may represent unique investing style. Search clusters exhibit phenomena demonstrated in prior style investing literature (Barberis and Shleifer 2003, Wahal and Yavuz 2013). Apart from comovement, we also show that past six-month cluster returns (search momentum) can be used to predict future one month return of member stock.

2.2 DATA AND SEARCH NETWORK

2.2.1 Main Data Sources

Our main data source is Yahoo! Finance. With an average monthly traffic of over 45 million visitors[3], it is one of the most popular investment portals among investors and it consistently ranks number one in terms of the popularity and the number of visitors.[4] Additionally, no other investment portal with a similar scale of visitors reveals the co-viewing pattern of investors.[5] 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 BAC (see Figure 1). Yahoo! computes the co-viewed data based on visitors' cookies[6] and uses a threshold to upload the most recent data to Yahoo! Finance.[7] We use the co-viewing data to identify subsets of stocks that attract investor attention during a certain time period.

---

3 http://www.ebizmba.com/articles/business-websites

4 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.

5 http://www.ebizmba.com/articles/business-websites

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

7 Co-viewed stocks are ranked based on their co-viewing frequency, and the top six co-viewed stocks are displayed to the user. We have separately verified the data generation process directly with the customer service at Yahoo! Finance.

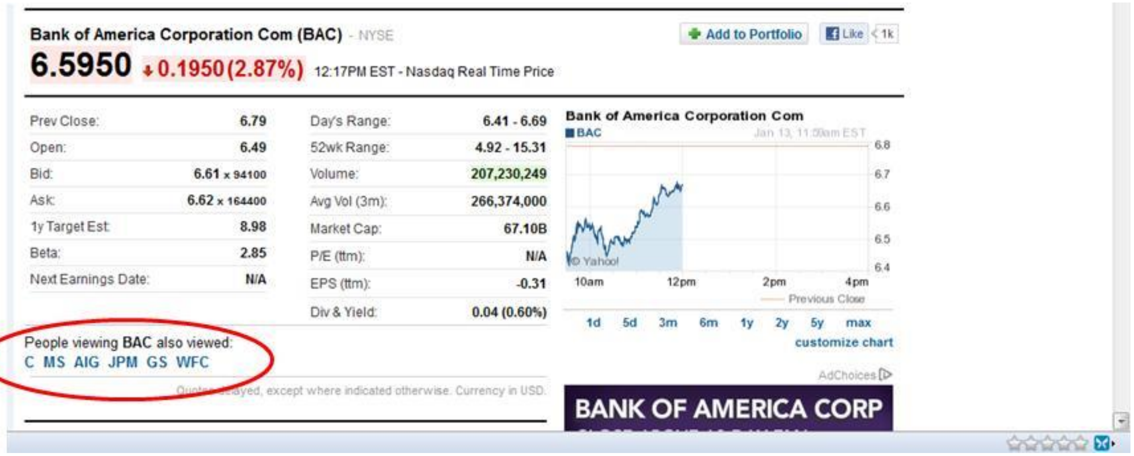

Figure 1: Example of Yahoo! Finance Co-Viewing Data

We focus on the search for stocks in the Russell 3000 index that account for 98% of the total market capitalization of all stocks trading in the U.S. This set is commonly used in the literature (Da et al. 2011, Diether et al. 2009, Evans et al. 2009, Haugen and Baker 1996). Using a Perl script, we collected daily co-viewing data for these stocks at 4pm CST every day during the period from September 15, 2011 to January 14, 2013.

In addition to Yahoo! Finance co-viewing data, we obtained the stock return data for each stock during the study period from the Center for Research on Security Prices (CRSP) database. We also collected daily news articles for our sample of stocks from Google News. Our final sample consisted of 2,900 stocks, as some stocks were delisted during the sample period.

2.2.2 Search Network

We use Yahoo! Finance co-viewed data in September 2011[8] to identify search clusters. In order to do this, we first map the co-viewed data into a directed graph where the nodes represent stocks and directed edge from a node to every other stock in the frequently co-viewed list. Each directed edge has a weight that is equal to the number of days the co-viewing relationship persists. An edge is bi-directional[9] if the two stocks are in each other's co-viewed data set. It is possible that a popular stock may appear in the co-viewed data of many stocks. However, many of these stocks may not appear in the co-viewed data of the popular stock. As we are interested in determining groups of stocks with common user interest, which show return correlation, we consider only those stocks that have bidirectional edges. This would indicate that investors pay more attention to certain subset of stocks as compared to others. In order to capture this difference in search intensity across different groups of stocks, we use a threshold co-viewing frequency to qualify bidirectional edges. We consider only those bidirectional edges where stocks appear in each other's co-viewing data for 2 weeks.[10]

Using the bidirectional edges, we determine *cliques*, which are the most basic units of search clusters. In graph theory, a clique is a subgraph with at least three

---

8 Because we started to collect data in the middle of September, the exact time period of September 2011 refers to September 15, 2011 to October 14, 2011. Similarly, October 2011 refers to October 15, 2011 to November 14, 2011. We use this month annotation throughout the paper.

9 Each bi-directional edge consists of two directed edges and thus has two weights. When we filter some non-frequently co-viewed stocks, we consider both weights of a bidirectional edge.

10 We filter out edges with lowest weight less than 2 weeks. We also try other filtering weights. Using lower weights increases the number of stocks that are classified in a cluster and using higher weights decreases the number of candidate stocks. However, stocks are always classified in the same cluster. Also, our qualitative results do not change by the change in the number of stocks associated with a search cluster.

members in which all members have an edge with each other (Luce and Perry 1949). We define a clique to be a subgraph of at least three stocks in which all members have *bidirectional* edges connecting each other. Consider $G_k(V_k, A_k, w_k)$ as a subgraph, where G_k is a subgraph, V_k is a set of vertices, A_k is a set of directed edges ($xy \in A_k$ implies that there is a directed edge from vertex x to vertex y), and w_k is a set of weights of directed edges. Thus, G_k is a clique if and only if $\forall u, v \in V_k, uv \in A_k, vu \in A_k$.

A maximal clique would represent a group of stocks where every stock has appeared in every other stock's co-viewed data as a large volume of users are interested in all these stocks. It is possible that some stocks belong to multiple maximal cliques. This would represent a scenario where different subsets of users focus on a different subset of overlapping stocks. These stocks may still share similar characteristics, but users may be focusing only on a subset of stocks due to limited cognitive resources. It is also an artifact of Yahoo! Finance data that the website gives only 6 frequently co-viewed stocks. To account for this possibility, we define a search cluster as a collection of all overlapping maximal cliques. For example, in Figure 2, there are two maximal cliques. The first maximal clique is formed by FE, AEP, and ECX, while the second maximal clique is formed by AEP, ECX, SO, ED, and DUK. The two maximal cliques have two overlapping members: AEP and ECX. We form a cluster by joining the cliques with overlapping members together. It is similar to hierarchical clustering techniques such as single-link clustering and complete link clustering that group stocks with smallest distance together. Thus, in Figure 2, FE, AEP, ECX, SO, ED, and DUK would be a 6-member cluster.

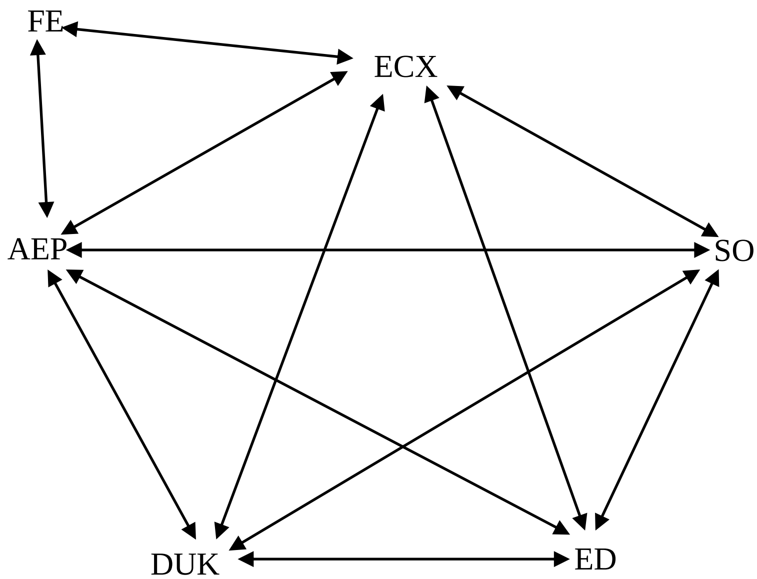

Figure 2: Cluster Formation

Based on the data collected in September 2011, we use network analysis software Pajek (De Nooy et al. 2005) to identify 79 search clusters that contain a total of 349 stocks.[11] Among these 79 clusters, there are 55 clusters with non-overlapping maximal cliques[12] and 24 clusters with overlapping maximal cliques[13]. The 349 stocks in these clusters have a total market capitalization of $6.61 trillion (as of October 14, 2011), representing approximately one-third of total market capitalization of all stocks traded on

[11] There are 5 stocks listed in the cluster identification period but delisted in the subsequent period of comovement analysis. In the comovement analysis, we remove these stocks. However, to prevent survivor bias, we also include the 5 stocks using available data up until the last date of listing as a robustness check. The results are qualitatively similar to our reported results.

[12] Among the 55 non-overlapping maximal cliques, there are three 7-member cliques, one 5-member clique, seven 4-member cliques, and forty-four 3-member cliques.

[13] The 24 clusters consist of 90 overlapping maximal cliques: four 5-member cliques, fifteen 4-member cliques, and seventy-one 3-member cliques.

the U.S stock exchanges. A large fraction of the stocks (i.e., 2,551) does not belong to any clique since there is no strong pattern of sustained co-viewing of stocks.

2.3 Model and Results

In this section, we present our main empirical results. Our key objective is to investigate whether search clusters represent investment habitats (i.e., stocks associated with a search cluster in the search network show excess return correlation or return comovement). Our empirical approach is as follows:

a) We first determine if the return correlation exists among stocks in a search cluster even after accounting for known common determinants of comovement such as market returns, news, and industry momentum.

b) Search clusters may be driven by the existing return correlation among stocks and may not reveal the investment habitats. In order to verify this, we consider changes to the search clusters such as addition and deletion of stocks from the search cluster and compare the return correlations of these stocks with their search cluster before and after the change. If the return correlation of a stock is not affected by its addition to a search cluster or deletion from a search cluster, it would suggest that the stocks in a search cluster may not represent an investment habitat.

We explain the related models and results below.

2.3.1 Measuring Comovement

To analyze the comovement of stocks within their respective clusters, we adopt the Barberis et al. (2005) method. Specifically, we estimate the following time series model:

$$\mathrm{R_{it}} = \beta_{0i} + \beta_{1i}\mathrm{R}_{\mathrm{C_i t}} + \beta_{2i}\mathrm{R}_{\mathrm{C_i t-1}} + \beta_{3i}\mathrm{R}_{\mathrm{S\&Pt}} + \beta_{4i}\mathrm{R}_{\mathrm{S\&Pt-1}} + \beta_{5i}\mathrm{FocalNews_{it}} + \beta_{6i}\mathrm{PeerNews}_{\mathrm{C_i t}} + \varepsilon_{\mathrm{it}}. \qquad (1)$$

Here, stock *i* belongs to cluster *C* at time *t* and R_{it} is the return of stock *i* on day *t*. $R_{C_i t}$ is the market-capitalization weighted return of cluster *C* of stock *i* at time *t*, excluding the return of stock *i*. To capture the effect of any lag in information diffusion across stocks, we include the lagged cluster return $R_{C_i t-1}$ in the model. This lagged term also accounts for the potential impact of search behavior that is influenced by the search cluster itself as investors are presented with other frequently viewed stocks based on past searches. Further, to account for the effects of the market, we include $R_{S\&Pt}$ in the model, which is the average market capitalization weighted return of S&P 500 at time *t*. We use a lagged term $R_{S\&Pt-1}$ to account for potential delayed effects of the market on the return of an individual stock. Control for market returns is similar to the control for environmental factors used by Oestreicher-Singer and Sundararajan (2012b) while analyzing demand correlations for books.

News can draw investor attention (Barber and Odean 2008) and in turn could influence the return comovement. For example, there could be positive news associated with the focal stock as well as other cluster stocks, which would lead investors to search and buy all the cluster stocks, resulting in stock comovement. So we control for the effect of news on return comovement. We define two news-related variables. $FocalNews_{it}$ is the log value of one plus the number of news articles that mention the focal stock *i* in the most recent 7 days (i.e., *t*, *t-1, …* *t-6*). This is similar to the method used in Da et al. (2011) to control for the effect of news. We also control for the effect of news of other stocks in a cluster on the focal stock. $PeerNews_{C_it}$ represents the log value of one plus the number of new articles that mention the peers in the same cluster as the focal stock *i* in the most recent 7 days. Panel A of Table 15 in Appendix A shows the summary statistics for all variables used in Model (1).

We use the estimation approach commonly used in finance (Boyer 2011, Chen et al. 2013, Da et al. 2011, Edgerton 2012) for asset returns. In order to capture the stock specific-effects, we estimate the regression models for each individual stock using OLS and report the average beta values. To account for the cross-sectional and temporal dependencies, we use two methods to measure the statistical significance of the coefficient estimates: (i) a parametric approach and (ii) a block bootstrapping method. In the parametric approach, we use asymptotic theory to determine robust standard errors. The details of this method are presented in Appendix B. Apart from the parametric approach, we compute the *p*-value using block bootstrap samples by randomly drawing a block of cross-sectional data 1,000 times[14] (see Appendix C).

[14] We also try block bootstrapping with 10,000 repeated random drawings. The results are qualitatively similar.

3.2 Baseline Results

We use the trading data from October and November 2011 to estimate Model (1). Note that we have defined the search clusters using search data from September 2011. This allows us to avoid any simultaneity bias in our estimates. Table 1 shows the estimation results for Model (1).

	$R_{C_i t}$	$R_{C_i t-1}$	$R_{S\&Pt}$	$R_{S\&Pt-1}$	$FocalNews_{it}$	$PeerNews_{C_i t}$
Mean	0.57	0.02	0.61	-0.01	0.0002	-0.0006
(Robust SE)	(0.03)	(0.03)	(0.05)	(0.05)	(0.0014)	(0.0009)
Bootstrap *p*-value	0.00	0.11	0.00	0.57	0.32	0.17
SD	0.48	0.37	0.71	0.49	0.01	0.01
Min	-1.53	-2.12	-2.56	-3.30	-0.05	-0.11
Max	2.56	3.61	3.44	3.13	0.10	0.04

Table 1: Comovement Regression Estimates: Baseline Model

We find that the average value of the search cluster coefficient $\bar{R}_{Ct}$ is 0.57, and it is significant at the 1% level in both parametric and block bootstrapping tests. This evidence indicates that the returns of stocks within the same search cluster are correlated. $\bar{R}_{S\&Pt}$ is positive and significant with a value of 0.61. This is consistent with prior studies on stock comovement (e.g., Barberis et al. 2005). Also, the lagged market return $\bar{R}_{S\&Pt-1}$ is insignificant, which suggests that the market prices incorporate information fairly quickly.

We also try different regression time periods (from 30 trading days to 50 trading days). The results remain qualitatively similar. Our cluster analysis period is different from the comovement estimation period and the cluster definitions can change over time.

If we were to use the precise definitions of clusters in the estimation period, the comovement of stocks within their respective clusters would have been even stronger. Thus, our current estimates represent the lower bound of cluster comovement.

3.3 Estimates Using Alternative Specifications

In this section, we extend Model (1) to account for the effects of other asset pricing factors—namely, size (SMB), book-to-market (HML), and momentum (UMD). Specifically, we estimate the following time-series model for each stock:

$$R_{it} = \beta_{0i} + \beta_{1i}R_{C_i t} + \beta_{2i}R_{C_i t-1} + \beta_{3i}MktRf_t + \beta_{4i}SMB_t + \beta_{5i}HML_t + \beta_{6i}UMD_t + \beta_{7i}FocalNews_{it} + \beta_{8i}PeerNews_{C_i t} + \varepsilon_{it}, \tag{2}$$

where $MktRf_t$ is value-weighted market portfolio return minus risk-free rate at time t obtained from CRSP, SMB_t is the return of the size factor at time t, HML_t is the return of the book-to-market factor at time t, and UMD_t is the return of the momentum factor at time t.[15] We estimate Model (2) using October and November 2011 stock data. As a robustness check, we replaced $MktRf_t$ with 49 Fama-French industry returns and got qualitatively similar results.

Panel B of Appendix Table 15 shows the summary statistics for the sample data. The estimation results reported in Table 2 show that even after including additional controls, the cluster coefficient $\bar{R}_{Ct}$ is significant and positive with a value of 0.48. The market factor $\overline{MktRf_t}$ has a slightly lower value of 0.43 and it is also positive and significant at the 1% level. In addition, the estimates of other three asset pricing factors

[15] Factor returns are obtained from http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

are significant. Similar to the results of Model (1), the lagged coefficient estimates and news variables are insignificant.[16]

For robustness, we also use equally weighted cluster returns instead of value-weighted cluster returns and replace weekly news with daily news in Models (1) and (2). In both instances, our results are qualitatively similar to the results reported in Tables 1 and 2.

[16] We consider a coefficient to be significant if the *p*-value is less than 0.05 and the test statistics of estimated beta divided by robust standard error is above 1.645.

	$R_{C_i t}$	$R_{C_i t-1}$	$R_{mt} - R_{ft}$	SMB_t	HML_t	UMD_t	$FocalNews_{it}$	$PeerNews_{C_i t}$
Mean	0.48	0.01	0.43	0.31	-0.15	-0.39	0.0003	-0.0002
(Robust SE)	(0.03)	(0.02)	(0.07)	(0.12)	(0.13)	(0.12)	(0.0013)	(0.0009)
Bootstrap *p*-value	0.00	0.05	0.00	0.00	0.01	0.00	0.36	0.33
SD	0.52	0.16	0.85	1.00	1.29	1.42	0.01	0.01
Min	-1.84	-0.64	-3.61	-2.98	-6.21	-12.63	-0.05	-0.11
Max	2.10	0.62	3.86	5.93	5.42	4.88	0.11	0.04

Table 2: Comovement Regression Estimates: Extended Model

3.4 Impact of Yahoo! on Stock Cluster Formation and Return Correlation

A potential endogeneity concern is that Yahoo! may influence the stock returns and may bias our results. One way this can happen is if Yahoo! is actively recommending stocks to investors.[17] To the best of our knowledge, Yahoo! was not using a recommendation engine for showing stocks except for displaying the co-viewed list of stocks based on a simple past co-viewing data. It is also possible that investors may click on co-viewed stocks presented to them and buy these stocks. However, our analysis shows that only 12% of the stocks actually form search clusters. This suggests that the co-listing of "also-viewed" stocks may not have a significant impact on the search process. Since the initial co-viewing list is only provided after substantial search history, any subsequent anchoring has little effect. We further address endogeneity concerns due to Yahoo's display of information by using a suitable instrument for the cluster returns and re-estimating the models using a 2SLS approach. We find that the estimates of OLS and 2SLS are not significantly different, which suggests that stocks are not influencing each other due to co-listing. Details of this analysis are included in Appendix D. This also addresses any potential endogeneity issues due to the influence of information such as "Top Picks" on the stock returns.

3.5 Dynamic Clusters and Comovement Changes

Next, we study how changes in search clusters induced by shifts in attention are related to the comovement patterns within dynamic habitats. If search clusters represent

[17] Related research on product networks (Kim et al. 2011, Oestreicher-Singer and Sundararajan 2012b) based product networks on information displayed by Amazon.com, which is known to use a recommendation engine.

investment habitats, then any change in the cluster membership of stocks over time should also be reflected in their return comovements. In particular, if a stock enters a new cluster, its return should show stronger comovement with the new cluster as compared to the return comovement with its previous cluster. Likewise, when a stock exits its current cluster, its return comovement relative to the previous cluster return should be weaker. Alternatively, the return correlations could be driving the cluster formation. In that case, we won't see these expected changes in the return comovement with changes in the cluster membership.

To understand the relationship between changes in cluster membership and the comovement, we divide our sample into four time periods, t_1 to t_4. We use the first time period t_1 (September 2011) to determine search clusters. Then, in time period t_3 (December 2011), we re-extract the search clusters. Based on the composition of search clusters in different time period, we obtain a list of stocks that changed their cluster membership. We analyze the comovement of stocks with their clusters in t_2 (October and November 2011) and t_4 (January and February 2012).

There are three scenarios: (a) stock addition to a cluster, (b) stock deletion from a cluster, and (c) stock switch to a different cluster. In (a), stock that did not belong to a cluster at t_1 is added to the cluster at t_3. In (b), a stock that originally belonged to a cluster at t_1 is no longer in the same cluster at t_3. In (c), the stock that originally belonged to a cluster at t_1 leaves the cluster and joins another cluster at t_3.

Figure 3 provides an example to illustrate the possible cluster changes over time. During period t_1, there are three clusters, namely, G1, containing stocks R, S, and T; G2, containing stocks U, V, and W; and G3, containing stocks X, Y, and Z. Stock Q is not part of any clusters. During period t_3, there are three clusters, namely, H1 (Q, R, S, T, U), H2 (V, W), and H3 (Y, Z) and one independent stock (X). From t_1 to t_3, we observe three types of changes: addition, deletion, and switch. Stock Q, which is independent at t_1, is added to cluster H1 at t_3. Stock X, which is a member of G3 in t_1, is deleted from the cluster at time t_3. Finally, stock U switches its cluster membership from G2 in period t_1 to H1 in period t_3. Note that 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.

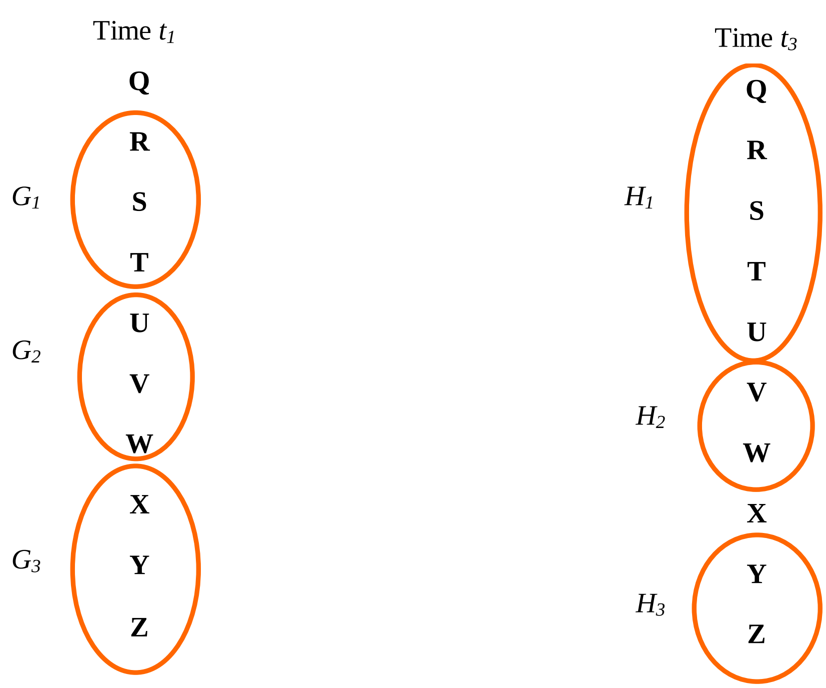

Figure 3: Examples of Clusters at Different Time Periods

In our data sample, between t_1 and t_3, there are 96 cluster additions, 205 cluster removals, and 34 cluster switches. Appendix Table 15, Panels C and D show the summary statistics for the three types of cluster changes in different time periods.

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

$$R_{it_2} = \beta_{0i}^{t_2} + \beta_{1i}^{t_2} R_{C_i t_2} + \beta_{2i}^{t_2} R_{C_i t_2 - 1} + \beta_{3i}^{t_2} R_{S\&P t_2} + \beta_{4i}^{t_2} R_{S\&P t_2 - 1} + \beta_{5i}^{t_2} FocalNews_{it_2} + \beta_{6i}^{t_2} PeerNews_{C_i t_2} + \varepsilon_{it_2} \quad (3A)$$

$$R_{it_4} = \beta_{0i}^{t_4} + \beta_{1i}^{t_4} R_{C_i t_4} + \beta_{2i}^{t_4} R_{C_i t_4 - 1} + \beta_{3i}^{t_4} R_{S\&P t_4} + \beta_{4i}^{t_4} R_{S\&P t_4 - 1} + \beta_{5i}^{t_4} FocalNews_{it_4} + \beta_{6i}^{t_4} PeerNews_{C_i t_4} + \varepsilon_{it_4} \quad (4A)$$

where R_{it_s} is the return of stock i in period t_s ; s = {2, 4}; R_{C_i, t_s} is the return of cluster C associated with stock i excluding the return of stock i

We also extend Models (3A) and (4A) by including Fama-French asset pricing factors in Models (3B) and (4B).

$$R_{it_2} = \beta_{0i}^{t_2} + \beta_{1i}^{t_2} R_{C_i t_2} + \beta_{2i}^{t_2} R_{C_i t_2 - 1} + \beta_{3i}^{t_2} MktRf_{t_2} + \beta_{4i}^{t_2} SMB_{t_2} + \beta_{5i}^{t_2} HML_{t_2} + \beta_{6i}^{t_2} UMD_{t_2} + \beta_{7i}^{t_2} FocalNews_{it_2} + \beta_{8i}^{t_2} PeerNews_{C_i t_2} + \varepsilon_{it_2} \quad (3B)$$

$$R_{it_4} = \beta_{0i}^{t_4} + \beta_{1i}^{t_4} R_{C_i t_4} + \beta_{2i}^{t_4} R_{C_i t_4 - 1} + \beta_{3i}^{t_4} MktRf_{t_4} + \beta_{4i}^{t_4} SMB_{t_4} + \beta_{5i}^{t_4} HML_{t_4} + \beta_{6i}^{t_4} UMD_{t_4} + \beta_{7i}^{t_4} FocalNews_{it_4} + \beta_{8i}^{t_4} PeerNews_{C_i t_4} + \varepsilon_{it_4} \quad (4B)$$

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. Following a cluster addition (deletion), stock i should show stronger (weaker) comovement with cluster C in period t_4 as compared to the comovement in period t_2. Therefore, $\overline{\Delta R}_{Ct}$ should be significant and positive (significant and negative). $\overline{\Delta R}_{Ct}$ is defined as $\overline{\Delta R}_{Ct} = \frac{1}{N}\sum_{i=1}^{N}\left(\beta_{1i}^{t_4} - \beta_{1i}^{t_2}\right)$, where N is the total number of stocks that experience cluster addition or deletion.

Likewise, we use the following model to analyze the impact of stocks switching to cluster H from cluster G by adapting the bivariate model proposed in Barberis et al. (2005).

$$R_{it_2} = \beta_{0i}^{t_2} + \beta_{1i}^{t_2} R_{G_i t_2} + \beta_{2i}^{t_2} R_{G_i t_2 - 1} + \beta_{3i}^{t_2} R_{H_i t_2} + \beta_{4i}^{t_2} R_{H_i t_2 - 1} + \beta_{5i}^{t_2} R_{S\&P t_2} + \beta_{6i}^{t_2} R_{S\&P t_2 - 1} + \beta_{7i}^{t_2} FocalNews_{it_2} + \beta_{8i}^{t_2} PeerNews_{G_i t_2} + \varepsilon_{it_2} \quad (5A)$$

$$R_{it_4} = \beta_{0i}^{t_4} + \beta_{1i}^{t_4} R_{G_i t_4} + \beta_{2i}^{t_4} R_{G_i t_4 - 1} + \beta_{3i}^{t_4} R_{H_i t_4} + \beta_{4i}^{t_4} R_{H_i t_4 - 1} + \beta_{5i}^{t_4} R_{S\&P t_4} + \beta_{6i}^{t_4} R_{S\&P t_4 - 1} + \beta_{7i}^{t_2} FocalNews_{it_4} + \beta_{8i}^{t_2} PeerNews_{H_i t_4} + \varepsilon_{it_4,} \quad (6A)$$

$$R_{it_2} = \beta_{0i}^{t_2} + \beta_{1i}^{t_2} R_{G_i t_2} + \beta_{2i}^{t_2} R_{G_i t_2 - 1} + \beta_{3i}^{t_2} R_{H_i t_2} + \beta_{4i}^{t_2} R_{H_i t_2 - 1} + \beta_{5i}^{t_2} MktRf_{t_2} + \beta_{6i}^{t_2} SMB_{t_2} + \beta_{7i}^{t_2} HML_{t_2} + \beta_{8i}^{t_2} UMD_{t_2} + \beta_{9i}^{t_2} FocalNews_{it_2} + \beta_{10i}^{t_2} PeerNews_{G_i t_2} + \varepsilon_{it_2} \quad (5B)$$

$$R_{it_4} = \beta_{0i}^{t_4} + \beta_{1i}^{t_4} R_{G_i t_4} + \beta_{2i}^{t_4} R_{G_i t_4 - 1} + \beta_{3i}^{t_4} R_{H_i t_4} + \beta_{4i}^{t_4} R_{H_i t_4 - 1} + \beta_{5i}^{t_4} MktRf_{t_4} + \beta_{6i}^{t_4} SMB_{t_4} + \beta_{7i}^{t_4} HML_{t_4} + \beta_{8i}^{t_4} UMD_{t_4} + \beta_{9i}^{t_2} FocalNews_{it_4} + \beta_{10i}^{t_2} PeerNews_{H_i t_4} + \varepsilon_{it_4,} \quad (6B)$$

where 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 a stock is expected to show comovement within its 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 $\overline{\Delta R}_{Ht}$ to be significant and positive, where $\overline{\Delta R}_{Gt} = \frac{1}{N}\sum_{i=1}^{N}(\beta_{1i}^{t_4} - \beta_{1i}^{t_2})$ and $\overline{\Delta R}_{Ht} = \frac{1}{N}\sum_{i=1}^{N}(\beta_{3i}^{t_4} - \beta_{3i}^{t_2})$.

Panel A of Table 3 shows the results for cluster additions. We find that $\overline{\Delta R}_{Ct}$ is positive and significant at the 1% level with a value of 0.28 using Models (3A) and (4A), and 0.22 using Models (3B) and (4B). This evidence supports our hypothesis that stocks show higher comovement with the new search cluster after the addition as compared to the comovement before the change. The estimate of $\overline{\Delta R}_{Ct-1}$ is insignificant, which is consistent with our findings in the previous subsection that lag value has limited power in explaining individual stock comovement. Other control variables are insignificant except $\overline{\Delta R}_{S\&Pt-1}$ and $\overline{MktRf}_t$, which are positive and significant at the 1% level in both parametric and block bootstrapping tests.

Next, Panel B of Table 3 shows the results for stock deletion from a cluster. As expected, $\overline{\Delta R}_{Ct}$ is negative and significant at the 1% level with a value of -0.10 in Models (3A) and (4A), and -0.07 in Models (3B) and (4B). The estimates imply that when a stock moves out of a cluster, its comovement with other stocks in the original cluster drops. Again, the lagged variable $\overline{\Delta R}_{Ct-1}$ and other control variables are insignificant except $\overline{\Delta R}_{S\&Pt}$ and $\overline{MktRf}_t$, which are positive and significant.

Last, Table 3, Panel C shows the results for stocks when they switch clusters. As expected, stocks switching from cluster G to cluster H show a lower comovement with cluster G after switching, as indicated by the negative and significant value of -0.31 (p value 0.00) for $\overline{\Delta R}_{Gt}$ for both baseline models (5A and 6A) and extended four-factor models (5B and 6B). These stocks show a higher comovement with cluster H as indicated by the positive and significant values of $\overline{\Delta R}_{Ht}$ (0.29 for baseline model and 0.3 for four-factor model). Thus, when a stock changes its cluster, it has stronger comovement with stocks in the new cluster than in the old cluster. The parameters of the lagged terms $\overline{\Delta R}_{Gt-1}$ and $\overline{\Delta R}_{Ht-1}$ and other controls are insignificant in either parametric or block bootstrapping tests.

Overall, these results provide evidence consistent with the habitat-based model of return comovement, which posits that as investment habitats shift, return comovement patterns among stocks change. Our evidence of attention-induced dynamic habitats indicates that investment habitats can change over time and generate time-varying comovement patterns in stock returns.

Panel A: Summary Statistics of Estimated Betas (Additions)

N=96	Model	$\Delta R_{C_i t}$	$\Delta R_{C_i t-1}$	$\Delta R_{S\&Pt}$	$\Delta R_{S\&Pt-1}$	$\Delta MktRf$	ΔSMB	ΔHML	ΔUMD	$\Delta FocalNews_{it}$	$\Delta PeerNews_{C_i t}$
Mean (SE)	A	0.28	-0.05	-0.31	0.21					0.0014	-0.0013
		(0.06)	(0.08)	(0.13)	(0.15)					(0.0012)	(0.0009)
	B	0.22	-0.04			-0.54	-0.07	0.15	-0.13	0.0014	-0.0002
		(0.06)	(0.03)			(0.14)	(0.18)	(0.22)	(0.18)	(0.0011)	(0.0009)
p-value	A	0.00	0.18	0.00	0.12					0.21	0.26
	B	0.00	0.11			0.00	0.48	0.23	0.29	0.22	0.43
SD	A	0.72	0.92	1.18	1.24					0.02	0.01
	B	0.83	0.30			1.54	1.89	2.12	1.79	0.02	0.01
Min	A	-2.13	-2.39	-3.83	-6.52					-0.09	-0.04
	B	-3.76	-0.78			-5.42	-7.38	-6.90	-9.01	-0.09	-0.04
Max	A	3.02	7.20	2.17	4.11					0.04	0.06
	B	3.27	0.81			3.40	9.45	8.02	6.94	0.05	0.06

Panel B: Summary Statistics of Estimated Betas (Deletions)

N=205	Model	$\Delta R_{C_i t}$	$\Delta R_{C_i t-1}$	$\Delta R_{S\&Pt}$	$\Delta R_{S\&Pt-1}$	$\Delta MktRf$	ΔSMB	ΔHML	ΔUMD	$\Delta FocalNews_{it}$	$\Delta PeerNews_{C_i t}$
Mean (SE)	A	-0.10	0.01	0.28	0.09					-0.0005	0.0004
		(0.03)	(0.03)	(0.08)	(0.09)					(0.0008)	(0.0006)
	B	-0.07	0.03			0.16	-0.09	-0.09	0.02	0.0006	0.0002
		(0.03)	(0.02)			(0.09)	(0.09)	(0.13)	(0.10)	(0.0008)	(0.0006)
p-value	A	0.00	0.34	0.00	0.11					0.27	0.26
	B	0.01	0.05			0.00	0.25	0.31	0.23	0.24	0.29
SD	A	0.58	0.53	1.22	1.02					0.01	0.01
	B	0.70	0.28			1.45	1.40	1.98	1.58	0.01	0.01
Min	A	-1.82	-4.01	-5.09	-2.42					-0.11	-0.04
	B	-2.73	-0.69			-5.46	-5.28	-11.36	-5.91	-0.12	-0.04
Max	A	2.94	1.93	7.13	4.56					0.04	0.11
	B	5.03	1.51			6.55	4.23	5.22	6.37	0.05	0.11

Table 3: Comovement Changes Around Cluster Additions, Deletions, and Switches

Panel C: Summary Statistics of Estimated Betas (Switches)

N=34	Model	$\Delta R_{G_i t}$	$\Delta R_{G_i t-1}$	$\Delta R_{H_i t}$	$\Delta R_{H_i t-1}$	$\Delta R_{S\&Pt}$	$\Delta R_{S\&Pt-1}$	$\Delta MktRf$	ΔSMB	ΔHML	ΔUMD	$\Delta FocalNews_{it}$	$\Delta PeerNews_{C_i t}$
Mean (SE)	A	-0.31	-0.18	0.29	-0.07	0.27	0.19					-0.0005	0.00004
		(0.08)	(0.10)	(0.09)	(0.09)	(0.16)	(0.17)					(0.0015)	(0.00119)
	B	-0.31	-0.09	0.30	-0.03			0.05	-0.26	0.24	-0.23	-0.0009	0.0007
		(0.08)	(0.05)	(0.09)	(0.08)			(0.16)	(0.17)	(0.22)	(0.17)	(0.0014)	(0.0012)
p-value	A	0.00	0.01	0.00	0.20	0.20	0.09					0.35	0.38
	B	0.00	0.04	0.00	0.38			0.46	0.11	0.25	0.07	0.29	0.24
SD	A	0.42	0.81	0.63	0.65	1.04	1.00					0.01	0.01
	B	0.53	0.47	0.75	0.50			1.24	1.29	1.13	1.12	0.01	0.01
Min	A	-1.29	-4.11	-0.66	-2.29	-2.03	-1.26					-0.02	-0.02
	B	-1.35	-1.98	-1.24	-1.48			-1.88	-3.01	-2.64	-2.25	-0.03	-0.01
Max	A	0.67	1.24	1.93	1.19	2.45	3.51					0.02	0.02
	B	0.77	0.41	1.89	1.34			2.76	3.16	2.58	3.30	0.01	0.02

Table 3: Comovement Changes Around Cluster Additions, Deletions, and Switches

(Continued)

2.4 ROBUSTNESS CHECKS

In this section, we present the results from a series of tests that examine the robustness of our main findings.

2.4.1 Tests Using Matching Stocks

In the first robustness test, we investigate the characteristics of stocks that belong to search clusters. Stocks in search clusters may exhibit similarities in characteristics such as size (Froot and Teo 2008), value (price-to-book ratio), and industry (Barberis and Shleifer 2003). In this scenario, search clusters we identify may exhibit similar levels of comovement as expected from these characteristics, which would suggest that investor attention follows shifts in firm fundamentals (e.g., size and value). If this is the case, then search may not reveal any additional information, as one could always derive the possible search clusters from the known characteristics.

To investigate the characteristics of stock clusters, we first determine how stocks within a cluster relate to each other in terms of size, value, and industry. Then, we conduct a placebo test to evaluate the comovement of matched stocks, which are similar to cluster stocks in terms of size, price-to-book ratio, and industry, but do not belong to the cluster. We compare the comovement of the original stocks within search clusters with those of matched stocks. We also identify other firm characteristics that could drive the comovement of stocks within the search clusters.

2.4.2 Measurement of Within-Cluster Similarity

To assess the level of similarity of firms in terms of size, value, and industry within each search cluster, we adapt the similarity index developed by Campbell et al. (1988). We first classify all CRSP stocks into decile based on their size (*s*) and price to book ratio (*v*). Then, we use the SIC codes to map each stock's industry to one of the 10 Fama-French industries.[18]

We compute the similarity index of cluster G along size and value dimensions as follows:

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

where D_A is the decile of stock A in size or value, $\overline{D}_G$ 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.

We compute the similarity index of cluster G in industry as follows:

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

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, or 0 otherwise, and N_G is the total number stocks in group G. If there is no mode, SI_G is 0.

[18] We also try the 49 Fama-French industries and find similar results. The difference in the average cluster similarity indices obtained using the 10 and 49 industries is small. Please see Table 4 Panel A for details.

All similarity indices fall in the range of [0, 1], where a value of 1 suggests that all stocks are the same in terms of the characteristic being considered, and a value of 0 suggests all members are different. Please refer to Appendix E for an example of search cluster similarity calculation.

Table 4 Panel A shows the sample statistics of similarity index in different attributes. Industry-based similarity estimates (FF10 and FF49) have the highest average similarity index, with a value of 0.88 (FF10) and 0.84 (FF49). It implies that most search cluster members are in the same industry. The average similarity index for size is the next highest with a value of 0.83, whereas the average similarity index for price-to-book ratio is slightly lower, with a value of 0.72.

Motivated by previous research demonstrating that stocks within the same geographic location show higher comovement (e.g., Kumar et al. 2013), we construct a similarity index based on the headquarters of firms in the same cluster. If the headquarters of all firms in the same cluster are in the same Census Bureau's Core Based Statistics Area (CBSA) code, the similarity index is 1. If all of them are in different state, the similarity index is 0. The location information is retrieved from COMPUSTAT and Compact Disclosure. As shown in Table 4 Panel A, the average similarity index of Region is not high, with a value of 0.32, which is far less than the average similarity indices based on size, value, FF10, and FF49. This evidence suggests that geographic proximity may not be a strong determinant of attention-induced stock clusters as other factors, since investors from different geographic areas search online and have access to similar information.

Panel A Similarity Index

Statistics	Size	Value	FF10	FF49	Volatility	Supply-chain	Region	Competitor
Mean	0.83	0.72	0.88	0.84	0.77	0.24	0.32	0.31
SD	0.15	0.20	0.21	0.26	0.13	0.34	0.34	0.41
Min	0.45	0.00	0.00	0.00	0.46	0.00	0.00	0.00
Max	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Panel B Dissimilarity Index

Statistics	Size	Value	FF10	FF49	Volatility	Supply-chain	Region	Competitor
Mean	0.43	0.56	0.14	0.17	0.56	0.76	0.68	0.69
SD	0.35	0.34	0.23	0.26	0.30	0.35	0.34	0.41
Min	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Max	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 4. Similarity and Dissimilarity Index Estimates

Some stocks may appear in the same search clusters because of the similarity in the stock symbol. Users are known to mistype ticker names in their stock price search, such as MCI and MCIC or AAPL and APPL (Rashes 2001). We manually check all cluster members to make sure that they do not have similar tickers to rule out ticker similarity as a potential cause of cluster formation.

We also construct a dissimilarity index for each attribute (e.g. size, value, industry). The idea is to find the mode in each attribute in each cluster and compute the proportion of members whose attributes are different from the mode. The dissimilarity index shows that some clusters are unique in different fundamental dimensions. As shown in Table 4 Panel B, most search clusters are dissimilar in supply-chain (0.76), followed by geographic location (0.68), value (0.56) and volatility (0.56). The dissimilarity index is lowest in industry with a mean of 0.14 in FF10 and 0.17 in FF49.

The results suggest that most cluster members are formed by industry. However, we do find clusters that are formed by firms in different industries, for example, Chipotle Mexican Grill (CMG), Green Mountain Coffee Roaster (GMCR), Priceline (PCLN) and Travelzoo (TZOO).

2.4.3 Placebo Tests

Since the similarity indices along various dimensions (size, value, and industry) are high, one may posit that the excess return comovement among cluster stocks is primarily due to similarities along these known dimensions. To evaluate the expected comovement generated by these stock characteristics, we implement a placebo test. Specifically, we compare the comovement of a stock in each cluster with the comovement of a placebo or matching stock that is not in the cluster. Placebo stocks match closely to stocks in the cluster along size, value, and industry dimensions. To construct the placebo cluster, we follow the approach described in Massa and Zhang (2009), who analyze the effect of style investing[19] on mergers.

First, for each stock in the search cluster under consideration, we find stocks that are in the same 10 Fama-French industry category. Then, we find the closest matched stock based on the size (i.e., market capitalization) and price-to-book ratio. These measures are computed one day before the comovement analysis (i.e., October 14, 2011). To find the closest matched stock, we compute the absolute differences between

[19] Style investing is an investment approach of investors who invest in stocks with certain style (e.g., big-cap/small-cap stocks, growth/value stocks, international/emerging market stocks). They may change their investment style based on the historical performance of stocks in the same category.

individual cluster member and each potential placebo stock using market capitalization and the price-to-book ratio.

Next, we rank all potential placebo stocks independently according to the absolute differences. We sum up the ranks of the two absolute differences in market capitalization and price-to-book ratio and choose the stock with the smallest sum as the final matched stock for each individual stock in the search cluster. If there are two or more stocks with equal sums, we choose the stock with the smaller absolute difference in market capitalization. If a matched stock turns out to be another stock of the same cluster, we ignore that stock and choose the next best matched stock. We also make sure that there is no overlap between the matched stocks for individual stocks of the search cluster under consideration. The summary statistics for matched stocks are shown in Panels A and B of Appendix Table 15.

After identifying the placebo stocks, we estimate Model (1), where the dependent variable is the return of the placebo stock $R_{P_i t}$. If positive comovement is caused by size, value, and industry, the beta of cluster return should be positive and significant. Table 5 shows the comovement estimates. As shown in Panel A, there is a positive comovement and the average beta is 0.14, which is significant at the 1% level in both statistical tests. Panel B shows that the average difference between the betas of cluster stocks and the placebo stock, which is 0.45, is positive and significant at the 1% level. These results imply that although comovement may be associated with size, value, and industry, cluster stocks possess features that cannot be fully explained by these three factors.

Consequently, there is a greater comovement among the cluster stocks as compared to the beta estimates of placebo stocks.

We also estimate the model using the four-factor model in Model (2) by replacing the cluster return by placebo return. As shown in Table 5, Panel C, there is positive comovement among placebo stocks and $\bar{R}_{Ct}$ is 0.04, which is insignificant in parametric test. Also, this magnitude is much smaller than the cluster stock comovement $\bar{R}_{Ct}$ of 0.48, which is significant in both tests (Table 2). As shown in Table 5, Panel D, the difference between the betas for the cluster stocks and the placebo stock is positive and significant at the 1% level with a value of 0.43.

Panel A: Summary Statistics of Estimated Betas (Placebo Stock Return as Dependent Variable) in Model 1

	R_{C_it}	R_{C_it-1}	$R_{S\&Pt}$	$R_{S\&Pt-1}$	$FocalNews_{it}$	$PeerNews_{C_it}$
Mean	0.14	-0.04	1.09	0.01	-0.0005	-0.0001
(SE)	(0.05)	(0.05)	(0.07)	(0.07)	(0.0016)	(0.0011)
p-value	0.00	0.00	0.00	0.39	0.20	0.55
SD	0.39	0.35	0.62	0.43	0.01	0.01
Min	-2.43	-1.78	-1.26	-1.99	-0.04	-0.04
Max	1.52	2.31	3.40	1.77	0.04	0.04

Panel B: Summary Statistics of Difference in Estimated Betas for Cluster Stocks and Placebo Stocks in Model 1

	R_{C_it}	R_{C_it-1}	$R_{S\&Pt}$	$R_{S\&Pt-1}$	$FocalNews_{it}$	$PeerNews_{C_it}$
Mean	0.43	0.06	-0.48	-0.01	0.0008	-0.0005
(SE)	(0.03)	(0.02)	(0.03)	(0.02)	(0.0013)	(0.0009)
p-value	0.00	0.00	0.00	0.28	0.19	0.15
SD	0.58	0.47	0.86	0.64	0.01	0.01
Min	-1.54	-2.25	-4.31	-3.28	-0.06	-0.09
Max	3.19	3.31	1.97	3.42	0.09	0.05

Panel C: Summary Statistics of Estimated Betas (Placebo Stock Return as Dependent Variable) in Model 2

	R_{C_it}	R_{C_it-1}	$R_{mt}-R_{ft}$	SMB_t	HML_t	UMD_t	$FocalNews_{it}$	$PeerNews_{C_it}$
Mean	0.04	-0.02	0.96	0.41	0.04	-0.18	-0.0003	0.0001
(SE)	(0.06)	(0.02)	(0.10)	(0.16)	(0.18)	(0.17)	(0.0016)	(0.0011)
p-value	0.00	0.99	0.00	0.00	0.17	1.00	0.33	0.46
SD	0.50	0.16	0.76	0.97	1.19	1.15	0.01	0.01
Min	-4.34	-0.96	-2.72	-3.31	-4.15	-3.71	-0.06	-0.04
Max	1.59	0.91	6.08	4.14	3.63	7.77	0.05	0.05

Table 5: Comovement Estimates from Placebo Tests

Panel D: Summary Statistics of Difference in Estimated Betas for Cluster Stocks and Placebo Stocks in Model 2

	$R_{C_i t}$	$R_{C_i t-1}$	$R_{mt} - R_{ft}$	SMB_t	HML_t	UMD_t	$FocalNews_{it}$	$PeerNews_{C_i t}$
Mean	0.43	0.04	-0.53	-0.10	-0.19	-0.21	0.0006	-0.0003
(SE)	(0.04)	(0.01)	(0.06)	(0.10)	(0.11)	(0.09)	(0.0013)	(0.0009)
p-value	0.00	0.00	0.00	0.08	0.02	0.00	0.31	0.33
SD	0.71	0.21	1.12	1.15	1.66	1.74	0.02	0.01
Min	-1.75	-0.73	-5.53	-4.03	-6.65	-11.58	-0.06	-0.09
Max	5.07	1.41	3.47	3.57	5.10	5.45	0.09	0.05

Table 5: Comovement Estimates from Placebo Tests

(Continued)

In short, the placebo test results show that fundamental factors, namely, size, value, and industry are contributing to the comovement among cluster stocks, but they cannot fully explain intra-cluster comovement. This evidence suggests that higher comovement within search clusters is likely to be associated with factors other than the known determinants of return comovement. It also suggests that stocks form a search cluster due to factors other than the known fundamental factors.

For robustness, we verify our results by creating placebo clusters using placebo stocks and testing the comovement of cluster stocks with the placebo cluster. We again find that the cluster stocks show higher comovement with cluster returns as compared to the placebo cluster returns.

We also investigate whether other characteristics such as volatility and supply-chain influence intra-cluster comovement. We do find that search cluster stocks share some of these characteristics. However, we find that even after controlling for these additional factors using appropriate matched stocks, the return comovement is stronger for cluster stocks. We report the results from this additional analysis in Appendix F.

2.4.4 Results Using an Extended Dataset

In the second robustness test, we evaluate the cluster comovement during a longer 15-month period from October 2011 to December 2012. As discussed earlier, since search clusters change with time, reflecting the changes in search patterns of investors, stocks can belong to different clusters at different points in time. So we compare the comovement of stocks with their original search clusters and new search clusters at different points in time during the 15-month period from October 2011 to December 2012. We consider search cluster definitions in September 2011, January 2012 and May 2012. We identify cluster stocks which switch to other search clusters in the subsequent months.

We determine the comovement of stocks with the original cluster and new cluster during the subsequent two months using baseline Models (3A) and (4A) and extended Fama-French four-factor Models (3B) and (4B). For example, if stocks move from cluster G to cluster H in October, we compare the comovement of these stocks with both clusters G and H in November and December.

Panels (A) to (C) in Table 6 show the average beta for original cluster returns and for the new cluster returns in different periods, and the average difference in betas. Stocks usually show a higher comovement with the most recent cluster they belong to as compared to the original cluster. The difference in betas is positive and significant in almost all cases. The results still suggest that clusters capture user attention over time and the stock search is associated with stock return comovement patterns. Further, changes in search patterns are associated with changes in the stock return comovements.

Panel A: Original Search Clusters Defined in September 2011

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

Table 6: Extended Comovement Analysis

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

Table 6: Extended Comovement Analysis

(Continued)

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

Table 6: Extended Comovement Analysis

(Continued)

2.4.5 Comention of News, Competitors and Pooled Regression

Comention of firms on news may catch attention of investors. Frequent exposure to comention stocks may cause investor to search some stocks together more often. We have re-run the main models in (1) and (2) with inclusion of comention news and the results remain qualitatively similar. Please see Appendix G for detail.

Furthermore, our search habitats may consist of commonsensical stocks with strong competitor relationship or strong fundamental similarity. The comovement pattern in search clusters may be caused by the commonsensical elements. To alleviate the concern associated with competitor relationship, we identify competitors from "Comparison" list in stock summary page of each focal stock and remove them from our analysis of comovement. To account for the similarity in fundamental similarity, we also control for placebo stock return in our main models. With all these measures, we still detect comovement phenomenon with removal of competitors. More discussion is available in Appendix H.

In our main analysis, we analyze comovement among individual stock rather than pooling. The primary reason is that pooling may violate homogeneity assumption in OLS regression (Bass and Wittink 1975). Departure from homogeneity may introduce bias in conventional significance tests (Wallace 1972) and distort conclusion of relationship among independent variables (Bass and Wittink 1975). However, we observe some previous studies (e.g., Froot and Teo 2008) also use pooled OLS regression with fixed effect. We repeat our analysis using same pooling strategy. We also try OLS regression with two-dimensional clustering by firm and date. The results are qualitatively similar. Please see Appendix I for detail.

2.5. Role of Attention in Search Cluster Formation

We show that stock clusters are dynamic (i.e., search clusters experience stock additions and deletions over time). Such changes may be associated with changes in fundamental values such as size and market value of stocks with time, which would lead investors to focus on different sets of stocks over time. However, at any given time there are several candidate stocks based on the fundamentals that could be part of the clusters, and only a few stocks actually end up in the search clusters. This would suggest that there are additional factors that drive the cluster formation and the stock comovement.

One such factor could be the change in attention for individual stocks. According to Barber and Odean (2008), attention influences the demand for stocks. One possibility is that while investors search a set of stocks based on their fundamental values, they focus on a subset, which grabs their attention. In order to test this, we investigate the role of attention in stock additions (including switches from some other clusters) to a cluster. If attention plays a role, the attention immediately before cluster addition should be significantly higher for the newly added stocks as compared to that for other candidate stocks. We identify other candidate stocks by matching the newly added stocks with similar fundamental values (i.e., similar size and value within the same Fama-French 10 industries). Note that these include stocks that are already in some other clusters.

We use the change in trading volume one month before cluster definition as a measure of change in attention. According to Barber and Odean (2008), extreme abnormal trading volume may catch the attention of investors. Retail investors (professional investors) are net buyers of stocks that experience the most positive (negative) abnormal trading volume (Barber and Odean 2008). In that case, if a stock experiences significant change in trading volume (positive or negative) in prior time period, it will catch the attention of investors and would be part of their search set as long as it also matches in other fundamental values with the existing stocks in the cluster. We estimate the probability of a stock being added to a search cluster as a function of its absolute change in the trading volume using the following logit model. Candidate stocks are newly added cluster stocks and all Russell 3,000 stocks that are not member of any existing clusters.

$$\Pr(Addition_{ict} = 1) = \beta_0 + \beta_1 \Delta Vol_{it-1} + \beta_2 DistMktCap_{ict-1} + \beta_3 DistP2B_{ict-1} + \beta_4 DissimFF10_{ict-1} + \beta_5 MktCap_{i,t-1} + \beta_6 P2B_{it-1} + \sum_{k=2}^{10} \beta_{7k} FF10_{ikt-1} + \varepsilon_{it} \quad (7)$$

where $Addition_{ict}$: binary variable that indicates whether stock *i* is added to cluster c at time *t*;

ΔVol_{it-1}: percentage volume change in the month before period t (in months);

$\Delta Vol_{it-1} = \left| \frac{\overline{Vol}_{i,t-2W:t-1W} - \overline{Vol}_{i,t-4W:t-3W}}{\overline{Vol}_{i,t-4W:t-3W}} \right|$, where $\overline{Vol}_{i,t-2W:t-1}$is the average trading volume of stock *i* in the most recent 2 weeks before time *t*; and $\overline{Vol}_{i,t-4:t-3}$is the average trading volume of stock *i* in the next most recent 2 weeks before time *t*.

$DistMktCap_{ict-1}$: Natural log of one plus absolute difference in market capitalization between stock *i* and average market capitalization of cluster *c* (excluding stock *i*) at time *t-1*

$DistP2B_{ict-1}$: Natural log of one plus absolute difference in price to book ration between stock *i* and average price to book ratio of cluster *c* (excluding stock *i*) at time *t-1*

$DissimFF10_{ict-1}$: The proportion of firms in cluster *c* at time *t-1* whose industry in FF10 is different from firm *i*

$MktCap_{i,t-1}$: Natural log of one plus market capitalization of firm *i* at time *t-1*

$P2B_{it-1}$: Natural log of one plus price to book ratio of firm *i* at time *t-1*

$FF10_{ikt-1}$: Fama-French 10-industry dummy variable to control for industry effect

We estimate the logit model using two-dimensional clustering technique that clusters standard errors by stock and time (Cameron et al. 2011, Thompson 2011). Table 7 shows the regression results. The coefficient for ΔVol_{it-1} is positive and significant (at 5%) with a value of 0.16. This suggests that a change in trading volume is likely to increase the odds a stock being added to a new search cluster. The coefficients for $DistMktCap_{ict-1}$ and $DistP2B_{ict-1}$are positive and significant suggesting that investors are more likely to include stocks whose market capitalization and price to book ratio in their search clusters. To the contrary, $DissimFF10_{ict-1}$ is negative and significant suggesting that investors tend to include stocks in the same Fama-French 10 industry as original search clusters. Therefore, the dissimilarity index in Fama-French 10 industry decreases the odds of a new stock being added to a search cluster. Other fundamental control variables (e.g., market value and price to book ratio) are significant and seem to suggest that big firms (high market value) and value stocks (low price to book ratio) are more likely to be added to a search cluster.

We also analyze the scenario of stock removal from a search cluster. A stock deletion suggests that investors co-search a stock less frequently with other cluster members. As investors have limited cognitive resources, they can focus only on a subset of stocks. We posit that the removed stocks do not catch much attention when compared with other cluster members, which may lead to its removal from the stock cluster. As before, we estimate the probability of stock removal as a function of the changes in the trading volume.

$$\Pr(Removal_{ict} = 1) = \beta_0 + \beta_1 \Delta Vol_{it-1} + \beta_2 DistMktCap_{ict-1} + \beta_3 DistP2B_{ict-1} + \beta_4 DissimFF10_{ict-1} + \beta_5 MktCap_{i,t-1} + \beta_6 P2B_{it-1} + \sum_{k=2}^{10} \beta_{7k} FF10_{ikt-1} + \varepsilon_{it} \quad (8)$$

where $Removal_{it}$: binary variable that indicates whether stock i is removed from cluster c at time t.

Candidate stocks for removal are all member stocks in existing search clusters. We estimate the logit model using two-dimensional clustering and summarize the regression results of the logit model in Table 7. We find that change in trading volume does not influence the likelihood of stock removal from a search cluster. To the contrary, the coefficients of $DistMktCap_{ict-1}$ and $DissimFF10_{ict-1}$ are both significant and positive suggesting that size and industry dissimilarity increase the odds of a stock being removed from a search cluster. Price to book ratio is the only significant fundamental control variable suggesting that growth stocks with high price to book ratio are more likely to be removed from an existing search clusters.

Thus, our results show that attention to individual stocks plays a role in search cluster formation and stock comovement. Our finding is unique, as the comovement literature has not considered the role of individual stock attention on the definition of investment habitats. Similarly, research on product networks has not explored the role of individual product attention on the network formation. Our results suggest that attention for an individual product can influence its association with other products.[20]

	Stock Addition (including switch)	Removal
ΔVol_{it}	0.16**	0.12
	(0.07)	(0.17)
$DistMktCap_{ict-1}$	0.56***	0.47***
	(0.17)	(0.10)
$DistP2B_{ict-1}$	1.62***	-0.19
	(0.17)	(0.18)
$DissimFF10_{ict-1}$	-2.15***	1.41***
	(0.31)	(0.49)
$MktCap_{i,t-1}$	0.65***	-0.09
	(0.13)	(0.07)
$P2B_{it-1}$	-1.33***	0.21**
	(0.26)	(0.09)
Constant	-12.57***	-5.29***
	(1.08)	(0.89)
	With FF10 Fixed Effect	With FF10 Fixed Effect
N	18,458	
		2,442
R^2	0.45	0.16

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

Table 7: Two-Dimensional Logit Regression Results

20 Note that we cannot expect negative attention to influence regular product search unless we have a setup where some customers may buy when there is reduced activity from other set of consumers.

2.6 SEARCH-BASED RETURN PREDICTABILITY

Search clusters may represent preferences of investors or a type of style investing. According to style investing literature, stocks that are in the same style exhibit both comovement and style momentum (Barberis and Shleifer 2003, Wahal and Yavuz 2013). Making use of style momentum, researchers find that it is possible to predict future return. Barberis and Shleifer (2003) suggest that a "hot" style may attract more inflows of investment to the style and thus generates momentum effect. We adapt the prediction model proposed by Wahal and Yavuz (2013) and test whether we can use past cluster return (excluding focal stock return) $R_{C_i t-6:t-1}$ to predict future focal stock return. We follow style investing literature to include controls such as lagged focal stock return $R_{i,t-6:t-1}$, industry momentum factor (Moskowitz and Grinblatt 1999), firm size measured by log average firm size in the previous six months, $ln\left(\overline{MktCap}_{i,t-6:t-1}\right)$ and value-growth ratio measured by log average price to book ratio in the previous six months, $ln\left(\overline{P2B}_{i,t-6:t-1}\right)$. Industry momentum is measured by lagged Fama-French 49 industry return ($FF49_{i,t-6:t-1}$). We also use search volume index, SVI, proposed by Da et al. (2011) to account for other unobserved effect. Log(SVI) is measured by log value of one plus Google Trends search volume of focal stock ticker or company name if the ticker is generic. We also try $\ln(ASVI_{i,t-6:t-1})$, which is $SVI_{i,t-6:t-1}$ minus log one plus median SVI 1 year before. We follow Wahal and Yavuz (2013)'s approach and analyze the data by Fama-MacBeth regression with Newey-West adjustment with a lag of 4 weeks. Fama-MacBeth regression is known to be good to control for cross-sectional correlation and

Newey-West adjustment can account for auto-correlation. The regression as shown in model (9) is rolling one week at a time. The dependent variable, $R_{i,t}$, is one month return of stock *i*.

$$R_{i,t} = \beta_0 + \beta_1 R_{C_i t-6:t-1} + \beta_2 R_{i,t-6:t-1} + \beta_3 FF49_{i,t-6:t-1} + \beta_4 ln\left(\overline{MktCap}_{i,t-6:t-1}\right) + \beta_5 ln\left(\overline{P2B}_{i,t-6:t-1}\right) + \beta_6 ln\left(SVI_{i,t-6:t-1}\right) + \varepsilon_{it} \quad (9)$$

Table 8 shows the prediction results. We tried prediction full sample using data from September 2011 to December 2012 and subsample without competitors. The coefficient of $R_{C_i t-1}$ is significant and positive all models. It shows that it is possible to predict stock return based on past quarter data. The coefficient of lagged focal stock return $R_{i,t-1}$ is also positive and significant which is consistent to prior study on style investing by Wahal and Yazul (2013). Lagged industry return is negative but insignificant in most cases, which imply that industry return may not be a good predictor in long-term prediction. Similarly, firm characteristics such as market capitalization and price to book ratio are all insignificant. Furthermore, SVI is found to be insignificant in our prediction model but ASVI is negative and significant at 10%. Da et al. (2011) find that Google Trends search volume can be used to predict short-term future returns of stock (1-2 weeks). Consistent to their findings, in our study, SVI cannot predict future returns of a longer horizon and ASVI captures price reversal.

	Full Sample	Full Sample	Without Competitors	Without Competitors
$R_{C_i t-1:t-6}$	0.038**	0.038**	0.041**	0.041**
	(0.017)	(0.017)	(0.018)	(0.018)
$R_{i,t-1:t-6}$	0.042***	0.043***	0.043***	0.045***
	(0.010)	(0.010)	(0.010)	(0.010)
$FF49_{i,t-1:t-6}$	-0.012	-0.008	-0.016*	-0.011
	(0.010)	(0.010)	(0.008)	(0.009)
$ln\left(\overline{MktCap}_{i,t-1:t-6}\right)$	0.003	0.003	0.003	0.003
	(0.003)	(0.003)	(0.003)	(0.003)
$ln\left(\overline{P2B}_{i,t-1:t-6}\right)$	-0.004	-0.003	-0.004	-0.003
	(0.012)	(0.011)	(0.013)	(0.012)
$ln(SVI_{i,i,t-1:t-6})$	0.001		0.001	
	(0.002)		(0.002)	
$ln(ASVI_{i,i,t-1:t-6})$		-0.015*		-0.015*
		(0.008)		(0.009)
Constant	-0.029	-0.028	-0.029	-0.028
	(0.049)	(0.049)	(0.051)	(0.051)
N	7,956	7,956	6,924	6,924
R^2	0.164	0.168	0.175	0.179

Table 8: Fama-MacBeth Regression of Future 1-month Return

2.7 Discussion and Conclusion

Collective search and search trends reveal powerful insights and provide the ability to predict outcomes. Hence, methods to understand, model, evaluate, and interpret collective search trends have become an important research topic. This study extends extant literature in the context of correlated searches often observed in many domains, including movies, product purchases, stocks, etc. This research is among the first studies to understand search behavior in the context of investing and using the correlated searches in understanding and predicting the underlying behaviors of assets (e.g., return comovement).

Analyzing stock search networks based on Yahoo! Finance data, we find that investors focus their attention on a few stocks at a time, resulting in unique search clusters. A well-known reason for focusing on limited set of stocks is due to cognitive limitation (Kahneman 1973, Li et al. 2013). Correlated searches occur for many reasons: intent to buy a set of stocks; monitor past investment choices; or simply curiosity about certain set of stocks. Despite heterogeneity in preferences, different groups of users focus on different sets of stocks, but certain dominant clusters—referred to here as search habitats—still form. Finance literature has investigated investment habitats using transactional-level data from various groups of investors. Habitats are known to form based on company size and institutional ownership (Froot and Teo 2008, Pindyck and Rotemberg 1993); indices such as S&P 500 (Barberis et al. 2005, Vijh 1994) and Nikkei 225 (Greenwood 2008); geography (Kumar et al. 2013, Pirinsky and Wang 2006); and volatility (Huberman and Dorn 2009) (see Table 4). Some investors may search for stocks that have buyer-supplier relationships. These habitats established in finance literature exhibit specific behavior; stock returns within these habitats tend to comove because of correlated trading (Barberis et al. 2005).

The search clusters or habitats may represent or suggest a proxy for systematic correlated trading. Such correlated trading may be lead to return comovement. Our results indeed show that stocks in the same search clusters show return comovement even after controlling for the fundamental factors known to result in comovement such as size, value, industry, supply-chain relationship, and volatility (Barberis et al. 2005). We do not suggest that these searches influence returns or that retail investors searching online influence the stock prices. However, correlated searches of a large of investors *reflect* the correlated investment choices made in the market that includes both retail and institutional investors. That is, even in the absence of investment choices or trading information of retail or instutional investors, the aggregated correlated searches show the underlying investment habitats. This is an exciting insight, since past research in finance primarily shows return comovements for different habitats (e.g., size) in isolation using ex-post-transactional data.

Information diffusion across stocks is a focus of substantial research. This study demonstrates that search cluster is a proxy for information diffusion as well. Further, correlated search may capture the effect of multiple factors that drive the investment habitats in the pre-choice mode and as a result can be used to detect changes in investment preferences without relying on transactional data.

The results show that search clusters are not constant; they evolve over time and the comovement patterns of underlying stocks change accordingly. This is an interesting finding, since past literature focusing on fundamentals are relatively stable. As discussed earlier, we show that one of the potential drivers of these changes in the cluster composition is the investor focus on a subset of stock with greater attention. Stock characteristics may change, which may cause investors to focus on different sets of stocks over time. Even then, many potential candidates for investment may meet investor preferences. In that case, investors may focus on stocks that are receiving higher attention in the market for various reasons—such as new products, partnerships, and earnings—and in turn creating awareness among the investors. This awareness drives the investor search process and leads to changes in the search cluster composition. The results are consistent with attention theory, which suggests that investors are net buyers of attention-grabbing stocks (Barber and Odean 2008). Our study extends this to show the role of investor attention in search cluster dynamics and time-varying comovement.

This study has important implications. Search clusters can reveal unusual or non-intuitive correlations such as multimarket presence within clusters. For instance, in our sample, we find that Abott Laboratories (NYSE: ABT), which operates in both personal health care and pharmaceutical markets, is in the same search cluster as personal health care companies such as Johnson & Johnson (NYSE: JNJ) and Procter & Gamble (NYSE: PG) and pharmaceutical companies such as Eli Lilly (NYSE: LLY) and Bristol Myers Squibb (NYSE: BMY). Similarly, corporate ownership can also be a factor driving the search cluster formation. For example, Xerox (NYSE: XRX) and Goodyear Tire & Rubber (NYSE: GT) have a representation on each other's boards and are part of the same search cluster. Kraft Foods (NYSE: KFT) is a spinoff of Altria (NYSE: MO), but both companies are part of the same search cluster despite being in different sectors since investors continue to hold interest in both companies

Previous research in investment habitats has primarily focused on fundamental habitats based on stock characteristics extracted from trading data. We provide a methodology to understand changes in the habitat using search history that can provide substantial input for understanding investment preferences and stock cross predictability on a real-time basis. Thus, our research contributes significantly to revealing investment interests and return comovement not identifiable for casual or professional investors. Further, as we use data from Yahoo! Finance, which is more likely to be used by retail investors, our research suggests that searches of a large number of retail investors can reveal outcome behaviors often perceived to originate from institutional investors.

This is similar to another growing area of IS research where collective sentiments can be used to predict returns or sales. The basic idea is that sentiments of a large number of users about stocks or products/services (e.g., movies, books) reflect market sentiment, and aggregated sentiments can be used as a predictor. In the context of investing, there is growing interest in building systems that capture collective sentiments, referred to as collectively intelligent systems (Watkins 2007), for helping investors make decisions. Several online firms such as Predictwallstreet.com, Stockpickr, Covestor, Gurufocus, and TickerSpy have emerged to extract collective sentiments. This study extends that research to correlated searches where aggregated searches can be used to determine stock comovement and cross predictability.

Our approach can be applied in other contexts to reveal correlated user preferences for different products or services. Making use of the revealed preferences, marketers can promote relevant products and information to online users and convert search to buying customers. For example, it is known that Mac users are more likely to stay in expensive hotels as compared to PC users.[21] This insight can be further improved by investigating the search correlation between different hotel brands and computer

[21] http://online.wsj.com/news/articles/SB10001424052702304458604577488822667325882

brands so that hotels can precisely target customers based on their correlated preferences for computers and hotels. Similarly, as firms create awareness about new products and lead to shifts in user attention, analysis of co-search can detect new associations between products, revealing changing user preferences that can be used by the firms for co-promotions and cross-predictions.

There are several limitations in our analysis that can be the basis for future research. While the Yahoo! Finance data reveals information about the co-viewing behavior, for every stock it shows only the top six co-viewed stocks. This may limit the size of clusters. Future research can explore other data sources such as Google Correlate and message boards to identify the precise composition of search clusters. Also, our dataset reveals only the search data, and then we make a connection between the search data and the return comovement. This analysis can be further improved if access to transactional data is also available for the same user base. It would also be useful to identify additional factors that drive investor attention and generate investment habitats. Last, it would be useful to compare the efficacy of different platforms in revealing search and investment patterns.

Chapter 3: Co-search Attention and Supply-Chain Relationship: The Case of Stock Return Predictability

3.1 INTRODUCTION

Internet enabled IT platforms such as portals, search engines, auction sites influence user search costs and have led to a number of research studies in IS focusing on the associated economic outcomes such as prices and market efficiency (Bakos 1997, Brynjolfsson and Smith 2000, Granados et al. 2012, Kuruzovich et al. 2008, Weber and Zheng 2007). IT platforms also capture user search which can reveal user preferences, for example, investment habitats in stock price search network (Leung et al. 2012). Search preferences for individual products or assets can be aggregated to reveal demand patterns for individual products or assets. There is much interest to build a strong foundation around search frequency, volume, and trends (e.g., Google search and trends) and market prediction. Wu and Brynjolfsson (2009) use search engine data to predict future housing market sales and prices. Choi and Varian (2012) show how to use Google Trends to forecast near-term values of economic indicators. More recently, Da et al. (2011) and Preis et al. (2013) also use Google Trends data to predict market movement. Search data may represent general interests of public in general and reflect the most up-to-date trends of individuals. Therefore, they are useful in predicting future demand.

Demand prediction is a popular topic in IS. IS researchers have primarily studied the impact of the changes in this search behavior of consumers and firms and the related product and pricing outcomes. Oestreicher-Singer and Sundararajan (2012b) use co-purchase network on Amazon to predict consumer demand on books. Granados et al.

(2012) investigate how Internet product information influences price elasticity in airline tickets. Kuruzovich et al. (2010) analyze how seller search behaviors influence the auction price of vehicle. Though prior research has demonstrated the demand predictability based on information networks, they do not investigate the underlying causes.

In this study, we try to bridge the literature gap and explain the cross-predictability across a group of correlated products from an information diffusion perspective. We analyze cross-predictability in the context of Internet finance portals. There are several reasons to focus on the context of stock investment. First, investors are sensitive to stock-related information. Stock price changes rapidly when related corporate information is available to the stock markets. Second, investors are active to search relevant investment information. Yahoo! Finance alone has over 45 million monthly visitors[22]. Search volume on stock related information is enormous. Therefore, stock market is an ideal place to test information diffusion. In this study, we investigate whether co-attention of a group of economically related stocks influences the speed of information diffusion, which in turn affects the performance of cross-prediction. When investors focus on a stock, they are likely to pay attention to a set of other stocks with different levels of economic linkages such as supply-chain relationship. When there is significant co-attention among these firms, economic shocks or impact to one party may influence other correlated firms immediately; that is, new information about supply-chain partners is quickly incorporated into the prices of focal stocks resulting in comovement of

[22] http://www.ebizmba.com/articles/business-websites

stock returns (Barberis et al. 2005). On the contrary, inattention to company linkages may provide opportunities for significant predictable returns in the future across its supply-chain partners (Cohen and Frazzini 2008). The inattention results in slow diffusion of information where the impact of customers will be propagated to suppliers some time later and reflected in suppliers' stock returns (Cohen and Frazzini 2008). The lagged responses thus contribute to cross-predictability.

Our core research question is how does retail investors' attention to economically related stocks influence information diffusion and cross-predictability of stock returns? We argue that the co-search frequency provides a proxy for the extent of co-attention among different economically linked stocks. If two stocks are frequently searched together, those stocks with high co-attention may exhibit high correlation in stock returns or return comovement (Leung et al. 2012). If two firms are economically linked, but do not show significant co-attention in search, then potentially the impact of one on the other may lag. The difference in the speed of information diffusion as a result of variation in co-attention intensity provides an opportunity to evaluate stock predictability.

Using supplier-chain relationship data from Bloomberg and co-search data from Yahoo! Finance, we analyze cross-firm predictability on a weekly basis from mid-September 2011 to December 31, 2012. Our results show that when supply-chain stocks are co-searched frequently, the lagged returns of the supply chain partners cannot predict the current returns of focal stocks precisely. However, in the absence of such co-attention there is significant cross predictability across supply-chain stocks. We can capitalize on the strong predictability by trading supply-chain related stocks with low co-attention intensity. Our simulated trading strategies show that we can earn average weekly returns of 35 basis points (annualized alpha of 20.77%).

Our research enriches the burgeoning body of literature on demand prediction. We are among the first to investigate the role of search attention in cross-predictability. In addition, previous research primarily focuses on network properties in product networks, for example, books (Dhar et al. Forthcoming, Oestreicher-Singer and Sundararajan 2012b), videos on YouTube (Susarla et al. 2012) and photos on Flickr (Zeng and Wei 2013). User attention and its impact on information diffusion are rarely investigated. In this research, we find that user attention significantly influence the speed of information diffusion. As user attention varies over time, future research on demand prediction should not assume user attention to be constant.

Furthermore, our study also contributes to finance research in the area of supply-chain cross prediction. Most prior studies in this area (e.g. Cohen and Frazzini 2008) primarily focus on prediction in industry level on a monthly basis. Our study is more granular focusing on firm level on a weekly basis. Apart from this, our simulated trading strategy shows that our prediction model, which incorporates investors' co-attention, works better than strategy considering pure supply-chain relationship strength (e.g., Menzly and Ozbas 2010). Our method of prediction may also complement existing prediction models based on sentiment analysis of daily news articles and message board postings (e.g., Antweiler and Frank 2004, Lam 2004, Tetlock 2007, Tsai and Hsiao 2010).

In addition, there is an increasing interest of IS research in the area of network economy. There is a call for more study on the distribution, diffusion, and inferential value of economic networks (Sundararajan et al. 2013). Our study shows that IT platform (e.g. Internet finance portal) can reveal aggregated attention or inattention of individuals on certain subjects (e.g. stocks). The attention (inattention) level may expedite (hinder) information diffusion in economic network. We can make inferential prediction based on different levels of attentiveness. The economic implications of our study are significant.

This paper contributes to building a strong foundation around online search and market outcomes, and enriches existing studies on information diffusion and cross-predictability among supply-chain stocks. It also showcases how big data analysis such as analysis of user co-search can improve the prediction accuracy for the market returns. As online activities now become more prevalent and transparent, researchers can access co-searching data (which are available in cookies, server log files and search queries) more easily and analyze in real-time rather than post-hoc evaluation. Our study may embark on new areas of research on information search and demand prediction.

3.2 Theoretical Background

Our study is grounded on attention theory. Search is an attention intensive process. When people search, they allocate a considerable amount of attention to certain aspects of an environment and ignore others (Li et al. 2013). In the search process, people spend time and effort to notice, interpret, and acquire information (James 1981, Kahneman 1973). Therefore, search is a process that requires enormous amount of cognitive resources (Li et al. 2013). Given the limited cognitive resources of human beings, it is found that online shoppers only search from a few places where they can collect information to meet their objectives in mind (Johnson et al. 2004). This is consistent to the theory of attention which posits that attention of individuals is limited and we cannot handle unlimited information (Kahneman 1973). Therefore, lack of attention may be a cause for slow information diffusion.

Information diffusion is a popular research topic in IS. Studying information diffusion in an email network, Aral et al. (Aral et al. 2007) found that functional relationship, strength of ties, demographics and network factors influence the volume and type of diffusion (vertical or lateral in organizational hierarchy). In a study of technology diffusion, it is found that interpersonal channels of communication is an important factor in making adoption decision (Brancheau and Wetherbe 1990). In another study of product diffusion on online social networks, it is found that both network characteristics and adopters' demographics influence adoption decisions (Katona et al. 2011). Apart from those factors, social interactions and geographic mobility have been studied in the diffusion and consumption of user generated content in video sharing network (Susarla et al. 2012) and mobile network (Ghose and Han 2011). Prior studies primarily focus on network factors and user characteristics that facilitate information diffusion and product adoption. User attention is usually assumed to be consistent over time. However, this assumption is not necessarily true in reality. It should be noted that information diffusion consists of two stages, namely, awareness and adoption, and awareness is a prerequisite of adoption (Kalish 1985). If individuals do not pay significant attention, they may not adopt a product/idea. Different from previous studies which investigate diffusion facilitators, we investigate inhibitors of information diffusion in this research. We posit that lack of attention is one of the inhibitors.

Lack of attention has been studied in stock market. Various theories have been proposed. Limited attention theory suggests that when there is scant investor attention to public accounting information, it may lead to stock mispricing (Hirshleifer and Teoh 2003). Investor distraction theory suggests that investor may under-react due to extraneous news and limited investor attention may cause market under-reactions (Hirshleifer et al. 2009). Previous research has also found that timing and outlets of information affect investors' attentiveness. Dellavigna and Pollet (2009) find that investors' attention is more diverted on Friday and their responses to earnings announcements on Friday are less vigorous than other weekdays. Huberman and Regev (2001) find that investors respond to the news of a cancer-curing drug by EntreMed more vigorously when it appears on New York Times than its earlier appearance in the academic journal *Nature*. Attention of investors to external information is not always consistent. Prior study has shown that many investors ignore important financial information in firms' financial statements (Hirshleifer et al. 2004). As a result, they may overlook some important information in their stock valuation.

With regard to online stock price search, prior study shows that stocks commonly searched form an investment habitat and members in the same habitat exhibit high similarity in returns correlation, which is also known as return comovement (Leung et al.

2012). Among a group of frequently searched stocks, searchers may recognize news happened to specific member stocks immediately. Information diffusion may be fast among frequently search stocks. As a result, member stocks in a tightly connected co-search network may exhibit high similarity in returns. Theories of comovement attribute such phenomenon to correlated fundamentals (e.g., size and market value), similar sentiments, and faster information diffusions (Barberis et al. 2005). Different from previous study, we primarily focus on slow information diffusion in this research. If investors do not pay high attention to economically linked stocks, information diffusion may be slow. They may not immediately react to events happened to those stocks. When they realize the information and take trading action, the reaction is lagged. As a result, a lead-lag pattern in stock returns may be exhibited among low co-attention partners. A partner firm first shows abrupt changes in stock price due to occurrences of some events. Due to slow information diffusion, a similar change in stock price of focal firms appears with some time delay. As a result, we can use the lagged returns of partner stocks to cross-predict the current returns of focal stocks.

3.3 METHODOLOGY

We build a cross-prediction model using lagged co-search intensity and lagged supply-chain partner returns. If cross-predictability exists, the lagged partner returns will be able to predict future returns of focal stocks. We explain explanatory and control variables in the following sections.

3.3.1 Attention Intensity

We use Yahoo! "also-viewed" data to measure attention intensity. Yahoo! Finance lists top six co-viewing stocks in each stock summary page. Figure 4 shows an example of Yahoo! Finance stock summary page. The circled area shows the top six "also-viewed" stocks[23]. When the majority of Yahoo! users who search stock A (e.g. AMD in Figure 4) also search B (e.g. NVDA in Figure 4), stock B is in the "also-viewed" list of stock A. Prior study has utilized Yahoo! co-attention data to analyze characteristics of online investment habitats (Leung et al. 2012). Beginning from mid-September 2011, we run a computer script daily to collect co-viewing data from Yahoo! Finance. We compute the average search intensity every week. If a stock appears in the "also-viewed" list of a focal stock in a week at least once, we assume that the co-attention between the focal stock and the "also-viewed" stock is high. Otherwise, we consider the co-attention to be low.

[23] The customer service of Yahoo! Finance confirmed that the top six stocks are the most frequently co-viewed stocks by online users who visit the current stock summary page.

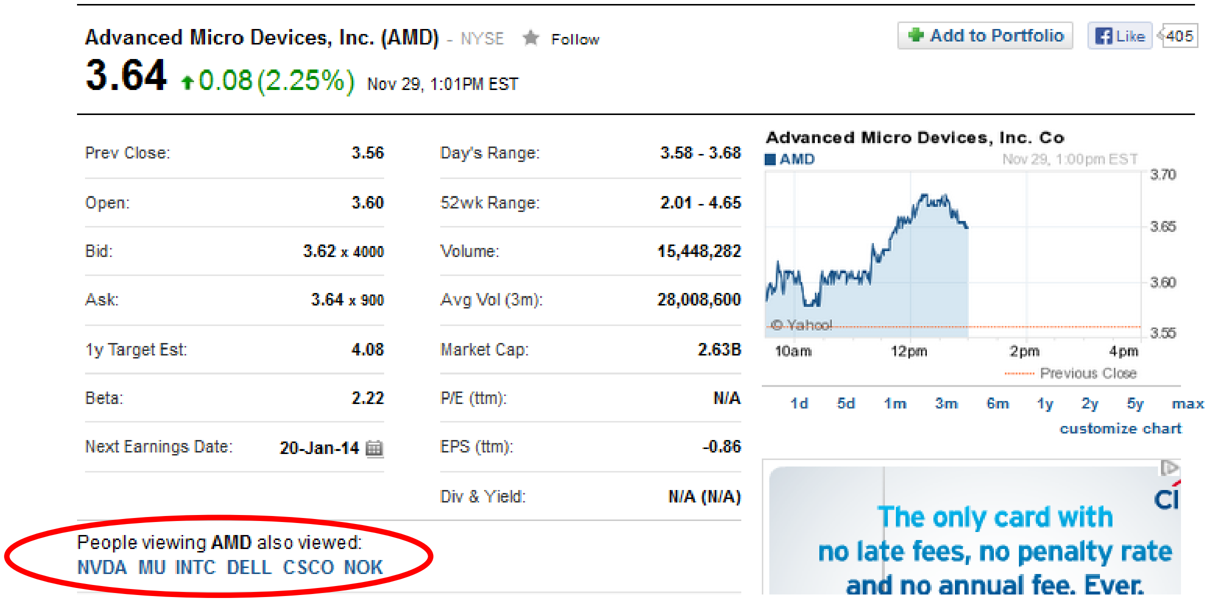

Figure 4: Example of Co-viewing Data in Yahoo! Finance

The use of Yahoo! Finance "also-viewed" list as a proxy for co-attention has several advantages over other measures used in prior finance research. Previously used measures of investor attention/inattention are imprecise. Cohen and Frazzini (2008) use mutual funds' joint holdings of supplier/customer stocks as proxies for investor attention. One limitation of the proxy data is that the authors assume the preference of mutual fund managers to be similar to individual investors. Also, the joint holding data are available only on a monthly basis and researchers assume the attention of investors does not vary too much within a month. Yahoo! Finance's co-searching data are available daily based on millions of Yahoo! users' search activities. Other conventional attention proxies include news, extreme past returns, trading volume (Barber and Odean 2008, Hou et al. 2008), and Google Trends search volume (Da et al. 2011). The primary limitation of those proxies is that they are firm-specific and do not account for the co-attention of investors across different stocks. Thus, they are not suitable in this study.

3.3.2 Supply-chain Relationship and Strength

Prior studies (e.g, Cohen and Frazzini 2008, Menzly and Ozbas 2010) have shown that supply-chain relationship may influence cross-predictability. Using Bloomberg Supply Chain Analysis (SPLC), we identify all stocks with supply-chain relationship to Russell 3000 stocks. SPLC classifies supply-chain partners into suppliers and customers and summarizes trading amount between focal stock and each supply-chain partners. The trading amount is based on the data reported by firms in their quarterly and annual earnings reports and industrial estimates by Bloomberg analysts. SPLC also provides data of revenue percentage and cost percentage, which are similar to Pandit et al. (2011)'s definition of dependency and exposure. In this study, we define dependency as the trading amount between a focal firm and a buyer divided by the total revenue of the focal firm and exposure as the trading amount between the focal firm and a customer divided by the total cost of goods sold of the focal firm. To account for different levels of supply-chain strength, we take into consideration of both dependency and exposure in the calculation of average partner returns.

The use of Bloomberg's dataset provides several advantages. Prior studies use Center for Research in Security Price (CRSP) Segment database to identify supplier and customer relationship and sales between two parties. However, CRSP Segment database only reports a small proportion of supplier-customer relationship data as Regulation Statement of Financial Accounting Standards (SFAS) No. 131 requires firms to disclose the identity of customers with more than 10% of total sales in quarterly reports (Cohen and Frazzini 2008). Also, customer names in the database are sometimes vague and researchers have to manually match the names to existing stocks in Compustat database which may result in some data loss (Pandit et al. 2011). Some studies (e.g. Menzly and Ozbas 2010) rely on Benchmark Input-Output Surveys of the Bureau of Economic Analysis (BEA) to identify the magnitude of trading between industries. Using the survey data, researchers can only identify supply-chain relationship among industries but not individual firms. As a result, cross-prediction in prior studies is restricted to intra-industry analysis. Furthermore, the survey is conducted once every 5 years by BEA and researchers assume that the industry supply-chain relationship does not change dramatically within 5 years. To overcome this limitation, we retrieve data from SPLC, which provides pairwise supply-chain data with the most recent trading amount between two firms. Furthermore, suppliers and customers are identified using Bloomberg tickers and it can alleviate the problems of manual company name matching.

3.4 Research Model

In this research, we conduct firm-level weekly return prediction. We focus on weekly analysis rather than monthly or daily analysis because weekly returns are more stable as shown in prior finance research (e.g., Hou 2007). Weekly return prediction is more precise than monthly return prediction. Also, we only have 66 weeks of data. The sample size may not be large enough for a monthly analysis. Furthermore, prior research has shown that predictions based on daily returns may suffer from microstructure influences, for example, bid-ask bounce and nonsynchronous trading (Hou 2007). Therefore, weekly return prediction seems reasonable.

We analyze the cross-predictability of supply-chain partners using model (1).

$$Ret_{i,t} = \beta_0 + \beta_1 Ret^{L}_{P_i,t-1} + \beta_2 Ret^{H}_{P_i,t-1} + \beta_3 Ret_{i,t-1} + \beta_4 MktRf_t + \beta_5 SMB_t + \beta_6 HML_t + \beta_7 MOM_t + \beta_8 Analyst_{i,t-1} + \beta_9 InstHldg_{i,t-1} + \beta_{10} News_{i,t} + \beta_{11} News_{i,t-1} + \beta_{12} CoNews^{L}_{P_i,t} + \beta_{13} CoNews^{L}_{P_i,t-1} + \beta_{14} CoNews^{H}_{P_i,t} + \beta_{15} CoNews^{H}_{P_i,t-1} + \varepsilon_{i,t} \qquad (1)$$

The dependent variable is focal firm *i*'s contemporary weekly return $R_{i,t}$. We follow prior finance research and use compounded daily return to compute weekly return (e.g., Hou 2007, Mech 1993, Rosenthal and Young 1990). Then, we compute supply-chain strength weighted partner returns with 1 week lag for each group of supply-chain partner $Ret_{P_i,t-1}$. If some supply-chain partners of a focal firm are listed in the co-searching list of the focal firm in any one day of previous week, they belong to high attention group (H). Otherwise, they are categorized as low attention group (L). We compute supply-chain strength (SC) weighted average partner returns separately for both attention groups. If a focal firm has both buyers and suppliers as its supply-chain partners, $Ret_{P_i,t-1}$ is defined as the average of dependency weighted buyer returns $R_{B_i,t-1}$ (Equation 2) and exposure weighted customer returns $R_{S_i,t-1}$ (Equation 3).

Menzly and Ozabas (2010) use the same approach to compute composite partner returns. If the focal firm only has one type of partners, the corresponding SC weighted average partner return (i.e. $R_{B_i,t-1}$ or $R_{S_i,t-1}$) is used as $R_{P_i,t-1}$. The main advantage of using a composite partner return is that it can reduce the number of parameters to be estimated while being model-justified (Menzly and Ozbas 2010). In our study, we do not care whether the type of supply-chain relationship (i.e. buyers and suppliers) has any differential impact on cross-predictability.

$$R_{B_i,t-1} = {\sum_{j\in B_i} Dep_{ij} \times R_{j,t-1}} \Big/ {\sum_{j\in B_i} Dep_{ij}} \quad (2)$$

$$R_{S_i,t-1} = {\sum_{j\in S_i} Exp_{ij} \times R_{j,t-1}} \Big/ {\sum_{j\in S_i} Exp_{ij}} \quad (3)$$

where Dep_{ij} is i's dependency on buyer j and Exp_{ij} is i's exposure to supplier j, and $R_{j,t-1}$ is weekly return of partner j at t-1. We compute the average supply-chain strength partner returns $R_{P_i,t-1}$ by taking the average of $R_{B_i,t-1}$ and $R_{S_i,t-1}$.

If there exists cross-predictability between lagged return of partner firms and current return of focal firms, we expect the coefficient of $R_{P_i,t-1}$ to be positive and significant.

We control for short-term reversal by including the lagged return of focal firm $R_{i,t-1}$ (Jegadeesh and Titman 1993, Menzly and Ozbas 2010). We also account for various market risk factors by controlling Fama-French 4 factors (i.e., MktRf, SMB, HML and MOM). Prior studies (e.g., Menzly and Ozbas 2010) show that attention can be affected by analyst coverage (Analyst) and institutional ownership (*InstHldg*). We also control for those factors. *Analyst* is measured by natural logarithm of (1 + the number of analysts following the focal stock). The data are retrieved from I/B/E/S database. *InstHldg* is calculated as natural logarithm of (1 + percentage of institutional holding). The data are obtained from Thomson Financial's 13F Holdings database. Furthermore, we control for news in our prediction model because news may capture investors' attention (Barber and Odean 2008) and has been used as a control in prior studies on predictability (e.g., Da et al. 2011). We control for news of focal firms by including news in contemporary week $News_{i,t}$ and news in previous week, $News_{i,t-1}$. News volume is calculated as the natural logarithm of one plus total number of news articles related to the focal firms. Da et al. (2011) uses similar approach in their calculation of news volume. We also control for comention news for each attention group in contemporary week (i.e., $CoNews^{L}_{P_i,t}$ and $CoNews^{H}_{P_i,t}$) and previous week (i.e., $CoNews^{L}_{P_i,t-1}$ and $CoNews^{H}_{P_i,t-1}$). The comention news is SC weighted average. We obtain news volume and comention news volume from all sources of news available in Factiva news database. We estimate our research model using two-dimensional clustering at firm and week level. Petersen (2009) shows that two-dimensional clustering produces better results for large panel data than OLS regression.

3.5 Research Results

We analyze co-search relationship of Russell 3,000 stocks and their supply-chain partners. As we only consider partners that are publicly listed in US stock exchanges, we remove some stocks from the analysis because they do not have any US listed partners. Furthermore, we remove small focal stocks with market capitalization less than 20th NYSE percentile because those thinly traded stocks are more volatile to market changes and may confound our cross-predictability results (Menzly and Ozbas 2010). Our full sample contains 102,912 firm-week data that comprise of 1,619 focal firms in 66 trading weeks. Table 9 shows summary statistics of our sample data and Table 10 shows correlation matrix. The correlation among independent variables is less than 0.70 in general except that contemporary news (comention news) and lagged news (comention news) has a correlation above 0.70. We are aware of the high correlation among news. However, the variation inflation factor (VIF) of our regression result is less than 5, which is below 10, the threshold of high multicollinearity. Nevertheless, we have also tried combining contemporary news and lagged news together and re-run the regression. The research findings are qualitatively similar.

Variable	N	Mean	SD	Min	Max
$Ret^{L}_{P_i,t-1}$	102,912	0.00	0.06	-0.74	1.49
$Ret^{H}_{P_i,t-1}$	102,912	0.00	0.03	-0.12	0.11
$Ret_{i,t-1}$	102,912	0.00	0.01	-0.11	0.10
$MktRf_t$	102,912	0.00	0.06	-0.74	1.49
SMB_t	102,912	0.00	0.02	-0.05	0.08
HML_t	102,912	0.00	0.01	-0.03	0.03
MOM_t	102,912	0.00	0.01	-0.02	0.02
$Analyst_{i,t-1}$	102,912	0.00	0.02	-0.04	0.03
$InstHldg_{i,t-1}$	102,912	1.66	0.78	0.00	3.91
$News_{i,t}$	102,912	0.56	0.13	0.00	1.20
$News_{i,t-1}$	102,912	2.44	1.39	0.00	8.64
$CoNews^{L}_{P_i,t}$	102,912	2.39	1.46	0.00	8.64
$CoNews^{L}_{P_i,t-1}$	102,912	0.11	0.34	0.00	5.42
$CoNews^{H}_{P_i,t}$	102,912	0.07	0.30	0.00	5.42
$CoNews^{H}_{P_i,t-1}$	102,912	0.13	0.53	0.00	6.06
$Ret^{L}_{P_i,t-1}$	102,912	0.11	0.51	0.00	6.06

Table 9: Summary Statistics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
(1) $Ret_{i,t}$	1	-0.05	-0.01	-0.06	0.54	0.31	-0.02	-0.37	-0.03	0.01	0.02	-0.01	0.01	0.00	0.00	0.00
(2) $Ret^{L}_{P_i,t-1}$	-0.05	1	0.40	0.56	-0.10	0.11	-0.16	0.00	-0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.00
(3) $Ret^{H}_{P_i,t-1}$	-0.01	0.46	1	0.23	-0.02	0.09	-0.08	-0.01	0.00	0.00	0.02	0.02	0.02	0.03	0.05	0.04
(4) $Ret_{i,t-1}$	-0.04	0.52	0.24	1	-0.07	0.04	-0.09	0.01	-0.01	0.01	0.01	0.02	0.01	0.01	0.00	0.00
(5) $MktRf_{t}$	0.49	-0.14	-0.05	-0.07	1	0.35	0.06	-0.65	-0.04	0.00	0.01	-0.02	0.02	0.01	0.00	0.00
(6) SMB_{t}	0.31	0.04	0.04	0.03	0.43	1	-0.31	-0.31	0.00	0.00	0.01	0.00	0.02	0.01	0.01	0.00
(7) HML_{t}	-0.02	-0.11	-0.09	-0.05	0.03	-0.30	1	-0.20	0.02	0.00	0.00	-0.01	-0.01	-0.01	0.00	0.00
(8) MOM_{t}	-0.34	0.05	0.04	0.02	-0.64	-0.34	-0.22	1	0.05	0.00	-0.01	0.03	-0.01	0.00	-0.01	0.00
(9) $Analyst_{i,t-1}$	-0.03	-0.02	-0.01	-0.01	-0.04	-0.01	0.02	0.05	1	0.09	0.20	0.24	0.13	0.14	0.09	0.09
(10) $InstHldg_{i,t-1}$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.12	1	-0.07	-0.07	-0.05	-0.08	-0.06	-0.08
(11) $News_{i,t}$	0.02	0.01	0.02	0.01	0.01	0.01	-0.01	-0.01	0.23	-0.04	1	0.77	0.47	0.39	0.32	0.30
(12) $News_{i,t-1}$	-0.01	0.01	0.02	0.02	-0.02	0.00	-0.01	0.02	0.26	-0.04	0.82	1	0.40	0.40	0.30	0.30
(13) $CoNews^{L}_{P_i,t}$	0.01	0.00	0.02	0.01	0.01	0.01	-0.01	-0.01	0.12	-0.04	0.41	0.35	1	0.71	0.38	0.36
(14) $CoNews^{L}_{P_i,t-1}$	0.00	0.00	0.02	0.01	0.00	0.00	-0.01	0.00	0.12	-0.04	0.34	0.35	0.77	1	0.39	0.42
(15) $CoNews^{H}_{P_i,t}$	0.00	0.00	0.04	0.00	0.00	0.01	0.00	-0.01	0.10	-0.05	0.38	0.36	0.39	0.38	1	0.84
(16) $CoNews^{H}_{P_i,t-1}$	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.11	-0.06	0.36	0.36	0.37	0.40	0.90	1

Upper Triangle: Spearman Correlation Matrix; Lower Triangle: Pearson's Correlation Matrix

Table 10: Correlation Matrix

As shown in Table 11, the estimated coefficient of lagged partner returns of low attention group is significant and positive. It supports our hypothesis that low attention leads to lagged market reaction and thus contributes to positive cross-predictability. To the contrary, the coefficient of lagged partner returns of high attention group is insignificant implying that there is no lagged market reaction among high attention partners. The lagged return of focal stocks is significant and negative. It is consistent to short-term reversion as documented in earlier finance research (Jegadeesh and Titman 1993). The Fama-French four factors are all significant except HML. Among the two attention variables, only institutional holding $InstHldg_{i,t-1}$ is significant and negative. It is consistent to the findings documented by Menzly and Ozabas (2010) that cross-predictability effects are weaker among stocks with high institutional holdings because there are more informed investors owning the stocks. With regard to news, only contemporary focal stock news is significant and positive. It shows that investors of focal stocks are aware of the news of stocks they invest in and take immediate reaction. The insignificant coefficient of lagged news reaffirms that investors react to focal firms' news immediately without any delay. The current and lagged comention news coefficients are insignificant implying that investors do not incorporate supply-chain partner news in their focal stock valuation.

Variables	Coefficients
$Ret^{L}_{P_i,t-1}$	0.0360***
	(0.0111)
$Ret^{H}_{P_i,t-1}$	0.0063
	(0.0300)
$Ret_{i,t-1}$	-0.0393***
	(0.0111)
$MktRf_t$	1.0673***
	(0.0160)
SMB_t	0.6220***
	(0.0271)
HML_t	-0.0118
	(0.0344)
MOM_t	-0.1051***
	(0.0251)
$Analyst_{i,t-1}$	0.0003
	(0.0004)
$InstHldg_{i,t-1}$	-0.0282**
	(0.0118)
$News_{i,t}$	0.0021***
	(0.0005)
$News_{i,t-1}$	-0.0005
	(0.0004)
$CoNews^{L}_{P_i,t}$	0.0010
	(0.0018)
$CoNews^{L}_{P_i,t-1}$	0.0002
	(0.0014)
$CoNews^{H}_{P_i,t}$	0.0001
	(0.0011)
$CoNews^{H}_{P_i,t-1}$	-0.0003
	(0.0010)
Constant	0.0092
	(0.0071)
	With FE
R^2	0.2689
N	102,912
Firms	1619
Weeks	66

Table 11: Cross Prediction Results

3.6 ROBUSTNESS TESTS

We observe a considerable number of focal firms have consistently low/high attention supply-chain partners throughout the research time period. Out of 1,619 focal firms, 796 firms have partners that do not change their attention intensity throughout the time period. As the number of firms is large, it may be the firm specific features of those attention unchanging stocks driving the cross predictability of supply-chain partner returns. To alleviate this concern, we compute the change frequency[24] of each focal firm and analyze a smaller sample of firms with different levels of change frequency: above 0%, 40% or above, 50% or above, and 60% or above. Using this approach, we guarantee that the attention level of some partners of a focal firm change some time in the research period. To account for partners that do not change over time, we add a new attention group, Others (or "O") in our research model. The group consists of partners that do not change their attention throughout the research time period. Lagged returns of partners belonging to group "O" ($Ret^{O}_{P_i,t-1}$), contemporary and lagged SC weighted commention news ($CoNews^{O}_{P_i,t}$ and $CoNews^{O}_{P_i,t-1}$) are included. The new model is shown in (4).

$$Ret_{i,t} = \beta_0 + \beta_1 Ret^{L}_{P_i,t-1} + \beta_2 Ret^{H}_{P_i,t-1} + \beta_3 Ret^{O}_{P_i,t-1} + \beta_4 Ret_{i,t-1} + \beta_5 MktRf_t + \beta_6 SMB_t + \beta_7 HML_t + \beta_8 MOM_t + \beta_9 Analyst_{i,t-1} + \beta_{10} InstHldg_{i,t-1} + \beta_{11} News_{i,t} + \beta_{12} News_{i,t-1} + \beta_{13} CoNews^{L}_{P_i,t} + \beta_{14} CoNews^{L}_{P_i,t-1} + \beta_{15} CoNews^{H}_{P_i,t} + \beta_{16} CoNews^{H}_{P_i,t-1} + \beta_{17} CoNews^{O}_{P_i,t-1} + \beta_{18} CoNews^{O}_{P_i,t-1} + \varepsilon_{i,t} \quad (4)$$

As shown in Table 12, lagged partner returns among low attention stocks has positive and significant cross predictability after removing firms whose partner attention do not change in the

[24] We compute SC weighted average partner attention for each focal stock weekly. Throughout the research time period, some focal firms may have the same SC weighted average partner attention because the majority of investors' co-attention intensity to supply-chain stocks does not change over time. To determine how frequent investors of a focal stock change their attention, we compute change frequency. It is defined as the ratio of the total number of weeks in which SC weighted average partner attention is different from the majority in the research time period.

research period. At the same time, control variables remain qualitatively similar to our main model. When change frequency increases, the cross predictability also increases accordingly. The results show that partner attention plays an important role in predicting stock cross predictability and such cross predictability is even stronger among firms whose investors shift attention frequently among supply-chain partners.

Variables	Change Freq > 0	Change Freq ≥ 0.4	Change Freq ≥ 0.5	Change Freq ≥ 0.6
$Ret^{L}_{P_i,t-1}$	0.0275***	0.0283**	0.0326*	0.0590***
	(0.0099)	(0.0124)	(0.0169)	(0.0202)
$Ret^{H}_{P_i,t-1}$	0.0265	0.0059	0.0257	0.0195
	(0.0186)	(0.0330)	(0.0251)	(0.0513)
$Ret^{O}_{P_i,t-1}$	-0.0130	-0.0219*	-0.0191	-0.0048
	(0.0104)	(0.0112)	(0.0134)	(0.0177)
$Ret_{i,t-1}$	-0.0351***	-0.0296**	-0.0444***	-0.0502***
	(0.0094)	(0.0146)	(0.0153)	(0.0187)
$MktRf_t$	1.0567***	1.0128***	0.9832***	0.9344***
	(0.0258)	(0.0389)	(0.0376)	(0.0341)
SMB_t	0.4820***	0.2839***	0.1939***	-0.0239
	(0.0488)	(0.0678)	(0.0681)	(0.0688)
HML_t	-0.1560**	-0.2967***	-0.3325***	-0.3179***
	(0.0636)	(0.0992)	(0.1014)	(0.0857)
MOM_t	-0.1771***	-0.2956***	-0.3017***	-0.2217***
	(0.0460)	(0.0677)	(0.0685)	(0.0768)
$Analyst_{i,t-1}$	0.0004	0.0005	0.0006	0.0007
	(0.0004)	(0.0005)	(0.0005)	(0.0005)
$InstHldg_{i,t-1}$	-0.0178	-0.0125	-0.0193	-0.0291
	(0.0124)	(0.0156)	(0.0138)	(0.0376)
$News_{i,t}$	0.0025***	0.0023**	0.0022*	0.0018
	(0.0006)	(0.0010)	(0.0013)	(0.0015)
$News_{i,t-1}$	-0.0002	-0.0003	0.0000	0.0010
	(0.0004)	(0.0008)	(0.0009)	(0.0009)
$CoNews^{L}_{P_i,t}$	0.0025	-0.0004	0.0011	0.0013
	(0.0020)	(0.0023)	(0.0024)	(0.0023)
$CoNews^{L}_{P_i,t-1}$	-0.0018	0.0006	-0.0017	-0.0017
	(0.0017)	(0.0021)	(0.0021)	(0.0020)
$CoNews^{H}_{P_i,t}$	-0.0003	-0.0001	0.0002	-0.0001
	(0.0013)	(0.0014)	(0.0014)	(0.0013)
$CoNews^{H}_{P_i,t-1}$	-0.0007	-0.0013	-0.0011	-0.0008
	(0.0012)	(0.0016)	(0.0015)	(0.0014)
$CoNews^{O}_{P_i,t}$	0.0000	-0.0014	-0.0021	-0.0037**
	(0.0022)	(0.0021)	(0.0026)	(0.0018)
$CoNews^{O}_{P_i,t-1}$	0.0018	0.0040**	0.0044***	0.0018
	(0.0018)	(0.0016)	(0.0017)	(0.0014)
Constant	0.0004	0.0018	0.0056	0.0103
	(0.0070)	(0.0084)	(0.0085)	(0.0229)
	With FE	With FE	With FE	With FE
R^2	0.2904	0.3094	0.3244	0.3231
N	53,176	17,284	11,768	5,634
Firms	823	267	180	87
Weeks	66	66	66	66

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

Table 12: Cross Predictability with Various Attention Change Percentage

Another concern of our research model is endogeneity. As Yahoo! Finance's co-searching list is open to the public, some investors may trade according to the list when they are frequently exposed to it. This conjecture is reasonable because Barber and Odean (2008) finds that individual investors are net buyers of attention grabbing stocks. If this is the case, the appearance of a stock on the co-searching list may be correlated to stock returns. As a result, the lagged returns of stocks in some attention groups may be endogenous. We try to solve the endogeneity issue using instrumental variable and run a 2SLS regression. We use Google Trends search volume data as instrumental variable. Da et al. (2011) find that weekly Google Trends search volume can be used to predict stock returns in the subsequent weeks. In other words, lagged Google Trends volume is correlated with contemporary stock returns. As the search volume data are collected from another search engine, it is not likely to be influenced by the co-searching list in Yahoo! Finance. Therefore, Google Trends search volume is likely to be a good instrumental variable. We follow Da et al. (2011)'s approach to identify Google Trends volume using individual stocks' ticker. If a ticker is too generic (e.g. EAT, LOOK), we use the company name as keywords to identity its Google Trends search volume.

In the first stage of 2SLS, we use stock returns of all supply-chain partners to run model (5). Then we obtain predicted stock return for individual partners and calculate predicted SC weighted partner returns in different attention groups using equations (2) and (3). In the second stage, we plug in the predicted SC weighted partner returns (e.g., $Ret^{L}_{P_i,t-1}$, $Ret^{H}_{P_i,t-1}$ and $Ret^{O}_{P_i,t-1}$) in models (1) and (4) and test cross-predictability of stocks in different attention groups.

$$Ret_{i,t} = \beta_0 + \beta_1 Trends_{i,t-1} + \beta_2 Ret_{i,t-1} + \beta_3 MktRf_t + \beta_4 SMB_t + \beta_5 HML_t + \beta_6 MOM_t + \beta_7 Analyst_{i,t-1} + \beta_8 InstHldg_{i,t-1} + \beta_9 News_{i,t} + \beta_{10} News_{i,t-1} + \varepsilon_{i,t} \quad (5)$$

As shown in Table 13, the first stage results show that our instrumental variable is positive and significant in explaining individual stock returns. In the second stage of 2SLS, we detect significant cross predictability using lagged partner returns of low attention group whereas the cross predictability of high attention group is insignificant. The Hausman test comparing estimates of main model (subsample with change frequency greater than 0) in Table 11 (Table 12) and 2SLS in Table 5 returns a chi-square test statistic of 0.07 (0.47). The insignificant test statistics imply that the estimates of both models are consistent and estimates in the main model (subsample) are more efficient.

Variables	First Stage	Second Stage (Full Sample)	Second Stage Change Freq > 0
$\widehat{Ret}^{L}_{P_i,t-1}$		0.0277**	0.0415**
		(0.0126)	(0.0194)
$\widehat{Ret}^{H}_{P_i,t-1}$		-0.0057	0.0193
		(0.0348)	(0.0336)
$\widehat{Ret}^{O}_{P_i,t-1}$			-0.0288
			(0.0211)
$Trends_{it-1}$	0.0025***		
	(0.0007)		
$Ret_{i,t-1}$	-0.0381***	-0.0357***	-0.0322**
	(0.0094)	(0.0118)	(0.0132)
$MktRf_t$	0.9595***	1.0672***	1.0580***
	(0.0390)	(0.0168)	(0.0257)
SMB_t	0.6569***	0.6240***	0.4865***
	(0.0508)	(0.0273)	(0.0487)
HML_t	-0.0695	-0.0214	-0.1601**
	(0.0841)	(0.0399)	(0.0645)
MOM_t	-0.1174***	-0.1030***	-0.1742***
	(0.0397)	(0.0270)	(0.0474)
$Analyst_{i,t-1}$	0.0002	0.0003	0.0004
	(0.0004)	(0.0004)	(0.0004)
$InstHldg_{i,t-1}$	-0.0114**	-0.0280**	-0.0180
	(0.0058)	(0.0118)	(0.0124)
$News_{i,t}$	0.0021***	0.0021***	0.0025***
	(0.0005)	(0.0005)	(0.0006)
$News_{i,t-1}$	-0.0008**	-0.0005	-0.0002
	(0.0004)	(0.0004)	(0.0004)
$CoNews^{L}_{P_i,t}$		0.0010	0.0025
		(0.0018)	(0.0019)
$CoNews^{L}_{P_i,t-1}$		0.0003	-0.0019
		(0.0014)	(0.0017)
$CoNews^{H}_{P_i,t}$		0.0001	-0.0003
		(0.0011)	(0.0013)
$CoNews^{H}_{P_i,t-1}$		-0.0003	-0.0007
		(0.0010)	(0.0012)
$CoNews^{O}_{P_i,t}$			0.0000
			(0.0022)
$CoNews^{O}_{P_i,t-1}$			0.0017
			(0.0018)
Constant	-0.0092	1.0672***	0.0005
	(0.0063)	(0.0168)	(0.0070)
	With FE	With FE	With FE
R^2	0.1643	0.2687	0.2901
N	157,030	102,912	53,176
Firms	2,498	1,619	823
Weeks	66	66	66

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

Table 13: Cross Predictability with Instrumental Variables

3.7 TRADING STRATEGY

As slow information diffusion occurs among low co-attention stocks, we can formulate a trading strategy to capitalize on the lagged information diffusion. Before the market starts, we sort all stocks according to their SC weighted partner returns in previous week. We group them into quintile (Q1: lowest and Q5: highest). Q5 (Q1) consists of stocks whose supply-chain partners have the most positive (negative) lagged returns. We form a value-weighted stock portfolio in each quintile. Due to slow information diffusion, positive (negative) news happened to partner stocks in previous week may lead to higher (lower) returns of focal firms in contemporary week. We can construct a trading strategy and gain a positive arbitrage by buying focal stocks whose partners have the most positive returns in previous week (i.e. Q5) and selling focal stocks whose partners have the most negative returns in previous week (i.e. Q1). The trading strategy is similar to the one adopted by Menzly and Ozbas (2010) who trade supply-chain related stocks without consideration of attention. We apply the aforementioned trading strategy to stocks with (1) low attention supply-chain partner stocks only, (2) high attention supply-chain partner stocks only, and (3) supply-chain partner stocks regardless of attention intensity. To avoid sample bias, we use out-of-sample data (Jan 1, 2013 to Dec 31, 2013) for the trading strategy.

Panel A Weekly Excess Returns on Value-Weighted Portfolios of Low Attention Supply Chain Related Firms

Strategies	Low (1)	(2)	(3)	(4)	High (5)	High - Low
Mean	0.0056	0.0079	0.0068	0.0084	0.0091	0.0035
SD	0.0160	0.0174	0.0170	0.0147	0.0149	0.0098
Sharpe Ratio	0.3499	0.4531	0.3997	0.5677	0.6113	0.3600
4-Factor	-0.0004	0.0006	-0.0001	0.0024***	0.0032***	0.0036**
Alpha	(0.0011)	(0.0009)	(0.0009)	(0.0008)	(0.0009)	(0.0015)

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

Panel B Weekly Excess Returns on Value-Weighted Portfolios of High Attention Supply Chain Related Firms

Strategies	Low (1)	(2)	(3)	(4)	High (5)	High - Low
Mean	0.0066	0.0056	0.0065	0.0064	0.0064	-0.0001
SD	0.0160	0.0148	0.0162	0.0145	0.0139	0.0084
Sharpe Ratio	0.4109	0.3814	0.4028	0.4404	0.4638	-0.0162
4-Factor	0.0002	0.0000	0.0002	0.0010	0.0009	0.0007
Alpha	(0.0010)	(0.0009)	(0.0008)	(0.0008)	(0.0008)	(0.0012)

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

Panel C Weekly Excess Returns on Value-Weighted Portfolios of Supply Chain Related Firms

Strategies	Low (1)	(2)	(3)	(4)	High (5)	High - Low
Mean	0.0057	0.0073	0.0066	0.0071	0.0078	0.0021
SD	0.0149	0.0156	0.0162	0.0146	0.0141	0.0079
Sharpe Ratio	0.3843	0.4682	0.4056	0.4875	0.5530	0.2640
4-factor	0.0000	0.0007	-0.0002	0.0013*	0.0021***	0.0021*
Alpha	(0.0008)	(0.0007)	(0.0006)	(0.0008)	(0.0006)	(0.0011)

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

Table 14: Trading Strategy

Table 14 shows the average weekly return of portfolios if we trade stocks in individual quintiles. Our trading strategy is to buy stocks in Q5 and sell stocks in Q1 (H-L). Panel A shows the results of trading only among low attention stocks. Q1 contains stocks whose partners experienced the most negative returns in the previous week and Q5 contains stocks whose partners experienced the most positive returns in the previous week. Due to slow information diffusion, our trading strategy shows increasing weekly return from Q1 to Q5. Our trading strategy (H-L) yields a mean weekly raw portfolio return of 35 basis points. To account for systematic market risks which may influence the raw portfolio returns, we compute portfolio alpha by running regression model (6).

$$R_{Port_{i},t} = \alpha + \beta_1 MktRf_t + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 MOM_t + \varepsilon_{i,t} \quad (6)$$

As shown in Panel A Table 14, the alpha of portfolio in Q1 is negative but insignificant and that of portfolios in Q5 is 32 basis points and significant at 5%. The insignificant alpha of Q1 may suggest that the focal stock investors do not incorporate the bad news of partner stocks immediately. As suggested by Hong et al. (2000), bad news diffuses slowly to the public. It may happen that it takes more than one week for investors to fully incorporate the bad news in their stock valuation. The portfolio of H-L generates the highest alpha of 36 basis points and significant at 5%. The annualized alpha of low attention (H-L) portfolio is 20.77% whereas that of baseline (H-L) portfolio is 11.51%.[25] The primary source of profits of our suggested portfolio comes from buying stocks in Q5.

Panel B shows the portfolio returns using only high attention stocks. Consistent to our earlier conjecture, the portfolio of H-L yields an insignificant alpha of 7 basis points or an annualized alpha of 3.68%. Furthermore, the portfolio returns between Q1 and Q5 are relatively similar. It seems to suggest that investors do not react to the lagged positive/negative returns occurred to supply-chain partners. This is consistent to the attention theory that increased attention may lead to faster information diffusion. Investors may have already incorporated the positive/negative news happened to supply-chain partners immediately. Therefore they do not show any lagged market reaction to the previous week's news.

[25] Annualized alpha is weekly alpha compounded over 52 weeks, $(1+\alpha)^{52}-1$.

Panel C shows the portfolio returns without consideration of attention. The weekly alpha is only 21 basis points and significant at 10%. The annualized alpha is only 11.51%. The results show that by trading supply-chain related stocks alone regardless of investor attention does not generate a high portfolio returns. The weekly alpha is much lower than that reported in Panel A. The results suggest that by incorporating attention intensity in our investment strategy, it is possible to earn higher portfolio returns.

3.8 DISCUSSION AND CONCLUSION

In this study, we show that we can use search network to reveal investors' co-attention intensity to supply-chain related stocks. The study contributes to the literature on information diffusion by providing empirical evidence support to limited attention theory in an investment context. When co-attention intensity is low between supply-chain stocks and focal stocks, information diffusion is slow. The slow information flow gives rise to the opportunity for cross-prediction. Our study may provide a direct response to Sundararajan et al. (2013) who calls for more research on network value. We contribute to the growing stream of research on network economy by demonstrating the inferential power and economic implications of stock co-search network.

Our study also contributes to finance research in return predictability. We are among the first to use co-search relationship in stock prediction. The new method may enrich existing research on stock prediction. Furthermore, in previous finance studies, cross-predictability of supply-chain stocks is attributed to information asymmetry. One source of information asymmetry is due to insufficient coverage of market segments by analysts (Menzly and Ozbas 2010). In this study, we show that another source of information asymmetry comes from lack of investors' attention. With control of analyst coverage and percentage institutional holdings, we find that cross-predictability only exists when investors' co-attention to supply-chain stocks is low. Though the Internet has reduced the search costs of investors, the limited cognitive resources of investors forbid them from searching unlimited stock information. As a result, investors do not always pay high attention to all supply-chain partners. This study also provides some empirical evidence to challenge market efficiency hypothesis. Our results seem to suggest that stock markets do not fully absorb supply-chain partner news if there is a lack of attention among investors.

Our study has important implications to practitioners. Investors may potentially gain from the lack of investors' attention to some supply-chain stocks. As shown in our out-of-sample simulated trading strategy, it is possible to make positive portfolio returns by trading on stocks whose investors pay less attention to their supply-chain partners. Furthermore, our results show that inattention is one of the major sources of slow information diffusion. To facilitate information diffusion, web designers may focus on web features (e.g. animation) that help in catching user attention. Online infomediaries may also provide often overlooked information to users based on their search interests.

Though our context of analysis is investment, our method may be applied to commodity products. Prior studies on market basket analysis show that we can use transactional data to identify product association. The association may be across product type (e.g. beer and diaper) and across product brand (e.g. Nike's running shoes and Apple's iPod). Given the knowledge of product association, we can use graph theory and network analysis to predict future demand of complementary products based on user search activities. As demand on one complementary product (e.g. shaving cream) is correlated to the demand on another item (e.g. razor), increase in search volume in one product may suggest more sales in both complementary products in near future. Our prediction model may be extended to product network and help retailers in predicting future demand. Better forecast of future demand can help in inventory control and shelf management. Our prediction model may also couple with sentiment analysis in social media to increase the accuracy in demand prediction.

Appendix

A. 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

Table 15: Summary Statistics

Panel C: Summary Statistics at t_2

Scenario	Addition				Deletion				Switch			
	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
Count (stock-days)												
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
$R_{C_i t_2}$	-0.00	0.04	-0.18	0.72	-0.00	0.03	-0.26	0.17	-0.00	0.03	-0.16	0.12
$R_{C_i t_2 - 1}$	-0.00	0.04	-0.18	0.72	+0.00	0.03	-0.26	0.24	-0.00	0.03	-0.16	0.12
$R_{S\&P t_2}$	-0.00	0.02	-0.04	0.04	-0.00	0.02	-0.04	0.04				
$R_{S\&P t_2 - 1}$	+0.00	0.02	-0.04	0.04	+0.00	0.02	-0.04	0.04				
$R_{G_i t_2}$									-0.00	0.03	-0.18	0.13
$R_{G_i t_2 - 1}$									-0.00	0.03	-0.18	0.13
$R_{H_i t_2}$									-0.00	0.02	-0.04	0.04
$R_{H_i t_2 - 1}$									+0.00	0.02	-0.04	0.04
R_{mt}-R_{ft}	-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
$PeerNews_{C_i t_2}$	2.22	2.15	0.00	6.32	3.15	1.79	0.00	6.55	3.58	1.73	0.00	6.32

Table 15: Summary Statistics

(Continued)

Panel D: Summary Statistics at t_4

Scenario	Addition				Deletion				Switch			
	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
Count (stock-days)												
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
R_{mt}-R_{ft}	+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
$PeerNews_{C_it_4}$	2.31	1.30	0.00	5.10	0.95	1.40	0.00	5.10	2.55	1.35	0.00	5.10

Table 15: Summary Statistics

(Continued)

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:

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

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 ≠ j is an element of the diagonal matrix of $\widehat{\Sigma_{\iota J}}$.

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, $i = 1, 2, \ldots, N$ is panel variable, and $t = 1, 2, \ldots, T$ 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 $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, L$;

Block 2 contains following data:

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

Block T-L+1 contains following data:

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

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 $\{\hat{\beta}^1, \hat{\beta}^2, \ldots, \hat{\beta}^{1,000}\}$.

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. 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). While the sentiments on message boards for a stock may be correlated with stock returns, these are not expected to influence the returns of other stocks in the cluster. 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. $Sentiment_{\mathrm{it}}$ of a stock *i* at time *t* can be expressed as

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

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 determine the potential endogeneity. In the first stage, we run an OLS regression on individual stocks as below:

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

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

We then compute estimated cluster return using the formula below.

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

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

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

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

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 16 and 17. As shown in Table 16, 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 $R_{S\&Pt}$ 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 $\hat{R}_{C_it}$ is significant and positive while the lagged cluster return $\hat{R}_{C_it-1}$ 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.

Panel A: Summary Statistics of Estimates in Model 7A

	$Sentiment_{it}$	$\overline{Sentiment}_{it-1,t-2}$	$R_{S\&Pt}$	$R_{S\&Pt-1}$	$FocalNews_{it}$	$PeerNews_{C_it}$
Coeff	0.0045	0.0037	1.35	0.02	0.0002	-0.0007
SE	0.0012	0.0016	(0.04)	(0.04)	(0.0007)	(0.0005)
p-value	0.00	0.03	0.00	0.13	0.38	0.22
SD	0.03	0.03	0.58	0.27	0.01	0.01
Min	-0.09	-0.17	0.26	-1.20	-0.04	-0.06
Max	0.26	0.13	3.05	1.88	0.08	0.03

Panel B: Summary Statistics of Estimates in Model 7B

	$Sentiment_{it}$	$\overline{Sentiment}_{it-1,t-2}$	$MktRf_t$	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
SE	0.0011	0.0013	(0.06)	(0.09)	(0.10)	(0.10)	0.0006	0.0005
p-value	0.01	0.05	0.00	0.00	0.01	0.00	0.34	0.57
SD	0.03	0.03	0.75	1.06	1.51	1.56	0.01	0.01
Min	-0.19	-0.17	-3.23	-2.97	-11.12	-9.56	-0.04	-0.06
Max	0.25	0.15	3.46	4.35	4.20	6.27	0.08	0.03

Table 16: Results for First Stage Regression

Panel A: Summary Statistics of Estimates in Model 8A

	$\hat{R}_{C_it}$	$\hat{R}_{C_it-1}$	$Sentiment_{it}$	$\overline{Sentiment}_{it-1,t-2}$	$R_{S\&Pt}$	$R_{S\&Pt-1}$	$FocalNews_{it}$	$PeerNews_{C_it}$
Coeff	0.79	-0.22	0.0047	0.0044	0.20	0.29	0.0001	-0.0002
SE	(0.13)	(0.12)	(0.0011)	(0.0016)	(0.16)	(0.14)	(0.0007)	(0.0006)
p-value	0.00	0.03	0.01	0.02	0.04	0.03	0.42	0.33
SD	1.96	1.09	0.03	0.03	2.45	1.50	0.01	0.01
Min	-14.07	-6.04	-0.08	-0.12	-9.84	-4.08	-0.04	-0.06
Max	7.74	4.05	0.26	0.22	14.91	8.84	0.08	0.03

Panel B: Summary Statistics of Estimates in Model 8B

	$\hat{R}_{C_it}$	$\hat{R}_{C_it-1}$	$Sentiment_{it}$	$\overline{Sentiment}_{it-1,t-2}$	$MktRf_t$	SMB_t	HML_t	UMD_t	$FocalNews_{it}$	$PeerNews_{C_it}$
Coeff	0.57	-0.01	0.0038	0.0028	0.19	0.31	-0.10	-0.32	0.0002	0.0002
SE	(0.11)	(0.02)	(0.0011)	(0.0013)	(0.13)	(0.11)	(0.15)	(0.15)	(0.0007)	(0.0005)
p-value	0.00	0.61	0.01	0.05	0.08	0.01	0.81	1.00	0.45	0.39
SD	2.05	0.31	0.03	0.03	1.96	1.34	2.41	2.37	0.01	0.01
Min	-17.00	-3.49	-0.20	-0.11	-8.26	-4.66	-21.94	-19.93	-0.05	-0.06
Max	7.36	0.91	0.26	0.13	7.97	5.77	7.77	8.72	0.08	0.04

Table 17: Results of Second Stage Regression

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 = $\left(I(D_{ARG} = 3) + I(D_{PX} = 3) + I(D_{APD} = 3)\right)\Big/_3 = {(0 + 1 + 1)}/_3 = 0.67;$

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* = ${}^2/_3 = 0.67.$

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 2009b). Some socioeconomic factors—for example, income, education level, occupation, ethnicity, and religion—may also help explain investors' preference for risky stocks (Kumar 2009b, 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. However, there are 28 clusters out of 79 that have a similarity of 0.6 and higher.

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 18 and 19 summarize the results for the coefficients of the cluster returns for the three different types of stocks.

Table 18 (Table 19) 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 20.

The coefficient for the cluster stock is positive and significant (Table 20). 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.

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
SD	0.48	0.40	0.40	0.59	0.55	0.43
Min	-1.53	-2.46	-2.46	-1.51	-1.74	-1.80
Max	2.56	1.57	2.21	3.22	3.22	1.78

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 (Robust SE)	0.48 (0.03)	0.05 (0.04)	0.08 (0.04)	0.43 (0.05)	0.39 (0.05)	-0.04 (0.03)
Bootstrap *p*-value	0.00	0.05	0.00	0.00	0.00	0.05
SD	0.52	0.48	0.46	0.70	0.63	0.49
Min	-1.84	-3.87	-3.08	-1.72	-1.94	-2.96
Max	2.10	1.63	2.56	4.61	4.03	1.97

Table 18: Comovement Comparison of Cluster Stock, Placebo Stock, and Volatility 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
SD	0.48	0.48	0.54	0.65	0.69	0.54
Min	-0.88	-1.26	-1.87	-1.10	-1.10	-1.91
Max	2.56	1.57	1.68	2.58	2.58	3.66

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 (Robust SE)	0.40 (0.05)	0.03 (0.06)	0.06 (0.06)	0.37 (0.07)	0.35 (0.07)	-0.02 (0.05)
Bootstrap *p*-value	0.00	0.29	0.13	0.00	0.00	0.43
SD	0.57	0.69	0.73	0.89	0.90	0.66
Min	-1.84	-4.25	-4.25	-1.61	-1.52	-1.87
Max	2.05	1.65	1.65	4.99	4.99	4.35

Table 19: Comovement Comparison of Cluster Stock, Placebo Stock, and Supply Chain Placebo

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 (Robust SE)	0.60 (0.05)	0.18 (0.05)	0.14 (0.05)	0.41 (0.06)	0.45 (0.07)	0.04 (0.06)
Bootstrap *p*-value	0.00	0.00	0.02	0.00	0.00	0.28
SD	0.47	0.41	0.46	0.58	0.53	0.59
Min	-0.26	-0.69	-1.62	-0.77	-0.66	-1.38
Max	2.56	1.32	1.35	2.58	1.87	1.40

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 (Robust SE)	0.44 (0.05)	0.08 (0.06)	0.13 (0.07)	0.36 (0.08)	0.31 (0.09)	-0.06 (0.08)
Bootstrap *p*-value	0.00	0.06	0.01	0.00	0.00	0.26
SD	0.55	0.50	0.67	0.76	0.84	0.75
Min	-0.95	-0.88	-1.56	-1.60	-3.11	-2.62
Max	2.05	1.57	3.71	2.92	1.89	1.43

Table 20: Comovement Comparison of Cluster Stock, Placebo Stock, and Supply Chain Placebo with Relationship Value

G. Comention of News

To control for the potential comention news effect, we adjust models (1) and (2) with inclusion of comention news. The comention 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 comention of both focal stock and partner stock in the same search cluster and compute a comention index, $CoNews_{C_i t} = \sum_{j \in C_i, j \neq i} MktCap_{jt} \times \ln(1 + News_{ijt}) \Big/ \sum_{j \in C_i, j \neq i} MktCap_{jt}$. $MktCap_{jt}$ is market capitalization of j at time t and $News_{ijt}$ is total number of news that mention both compnies 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 $FocalNews_{it}$. We analyze models (i) and (ii) as follows.

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

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

As shown in Tables 21 and 22, the main results do not change with inclusion of comention 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. Comention news is also found to be insignificant in both (i) and (ii)

	$R_{C_i t}$	$R_{C_i t-1}$	$R_{S\&Pt}$	$R_{S\&Pt-1}$	$FocalNews_{it}$	$CoNews_{C_i t}$
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
SD	0.52	0.43	0.72	0.52	0.01	0.01
Min	-1.89	-1.91	-2.34	-4.51	-0.02	-0.09
Max	2.45	4.85	3.81	2.82	0.07	0.05

Table 21: Comovement Regression with Comention News Volume: Baseline 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
SD	0.57	0.19	0.84	1.04	1.36	1.41	0.01	0.01
Min	-2.40	-1.81	-3.58	-2.84	-6.66	-11.17	-0.02	-0.11
Max	2.08	0.62	3.27	5.71	4.78	4.04	0.07	0.05

Table 22: Comovement Regression with Comention News Volume: Extended Model

H. 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 competitors in our analysis of comovement. We also include placebo cluster returns to account for potential influence due to commonsensical elements.

Competitors 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 competitors with transitive relationship. For example, if B is a competitor of A and C is a competitor of B, we consider A, B and C to be competitors though C does not appear in the "comparison" list of A. If a search cluster is formed by pure competitor 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 competitors 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 focal stock (please see 4.1.2 for detail) and compute placebo cluster return R_{PC_it}. 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 23 and 24, we still find positive and significant comovement with control of commonsensical elements. The baseline model shows that the magnitude of comovement is 0.56 whereas the extended model shows comovement with a magnitude of 0.48. Placebo stock return, R_{PC_it}, is positive and significant in both models with a value of 0.11 (baseline model) and -0.0001 (extended model). The variable is only significant in the baseline model. With control of Fama-French four factors, it become insignificant in the extended model. We also test the difference in coefficients between R_{C_it} and R_{PC_it}. Parametric and block bootstrapping tests both show that the difference is positive and significant suggesting the magnitude of comovement with search clusters is much higher than that with placebo clusters that are formed purely by similarity in fundamentals. The control variables are similar to previous analysis results.

	R_{C_it}	R_{C_it-1}	R_{PC_it}	$R_{S\&Pt}$	$R_{S\&Pt-1}$	$FocalNews_{it}$	$CoNews_{C_it}$
Mean	0.56	0.02	0.11	0.48	-0.01	0.0009	-0.0009
(Robust SE)	(0.03)	(0.02)	(0.04)	(0.07)	(0.05)	(0.0004)	(0.0010)
Bootstrap *p*-value	0.00	0.11	0.00	0.00	0.36	0.01	0.27
SD	0.49	0.42	0.53	0.88	0.51	0.01	0.01
Min	-1.44	-2.05	-5.08	-2.35	-4.76	-0.02	-0.08
Max	2.69	5.00	2.73	9.03	2.15	0.04	0.07

Table 23: Comovement Regression without Competitors: Baseline Model

	R_{C_it}	R_{C_it-1}	R_{PC_it}	$MktRf_t$	SMB_t	HML_t	UMD_t	$FocalNews_{it}$	$CoNews_{C_it}$
Mean	0.48	0.01	-0.0001	0.43	0.31	-0.14	-0.36	0.0004	-0.0009
(Robust SE)	(0.03)	(0.01)	(0.0357)	(0.08)	(0.11)	(0.12)	(0.11)	(0.0004)	(0.0010)
Bootstrap *p*-value	0.00	0.04	0.56	0.00	0.00	0.01	0.00	0.15	0.29
SD	0.53	0.16	0.53	0.87	1.02	1.39	1.33	0.01	0.01
Min	-2.20	-0.65	-4.64	-3.12	-3.64	-8.65	-9.54	-0.04	-0.11
Max	2.00	0.59	2.16	3.72	5.44	6.29	3.75	0.02	0.07

Table 24: Comovement Regression without Competitors: Extended Model

I. Pooled Regression Results

We repeat our main model and robustness tests (i) to (vi) using OLS regression with fixed effect and robust variance estimator. The results are shown in Table 25. Using pooled regression and with various controls, we still find positive and significant comovement among search clusters. Furthermore, we repeat the pooled regression with two dimensional clustering at firm and date level. The results remain qualitatively similar.

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

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

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

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

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

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

Model	(i)	(ii)	(iii)	(iv)	(v)	(vi)
$R_{C_i t}$	0.50*** (0.02)	0.48*** (0.02)	0.50*** (0.02)	0.48*** (0.02)	0.47*** (0.02)	0.46*** (0.02)
$R_{C_i t-1}$	-0.01 (0.02)	-0.00 (0.01)	-0.01 (0.02)	-0.00 (0.01)	-0.01 (0.02)	-0.00 (0.01)
$R_{PC_i t}$					0.18*** (0.02)	0.16*** (0.02)
$R_{S\&Pt}$	0.69*** (0.03)		0.69*** (0.03)		0.51*** (0.03)	
$R_{S\&Pt-1}$	0.02 (0.02)		0.02 (0.02)		0.02 (0.02)	
$MktRf_t$		0.45*** (0.04)		0.46*** (0.04)		0.32*** (0.04)
SMB_t		0.27*** (0.05)		0.26*** (0.05)		0.23*** (0.05)
HML_t		-0.14** (0.07)		-0.14** (0.07)		-0.13* (0.07)
UMD_t		-0.34*** (0.07)		-0.33*** (0.07)		-0.30*** (0.07)
$FocalNews_{it}$	-0.0000 (0.0005)	0.0000 (0.0005)	0.0006** (0.0003)	0.0005* (0.0003)	0.0006** (0.0003)	0.0005* (0.0003)
$PeerNews_{C_i t}$	-0.0002 (0.0004)	0.0001 (0.0004)				
$CoNews_{C_i t}$			+0.0000 (0.0005)	-0.0002 (0.0005)	-0.0000 (0.0005)	-0.0002 (0.0005)
Constant	-0.0029 (0.0020)	-0.0036* (0.0019)	-0.0068*** (0.0026)	-0.0057** (0.0025)	-0.0070*** (0.0025)	-0.0060** (0.0024)
	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
R2	0.43	0.44	0.43	0.44	0.44	0.44

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

Table 25: Pooled Regression Results

References

Aggarwal, R., Gopal, R., Gupta, A., Singh, H. (2012) Putting money where the mouths are: The relation between venture financing and electronic word-of-mouth. *Information Systems Research* 23(3-Part-2): 976-992.

Ansari, A., Essegaier, S., Kohli, R. (2000) Internet recommendation systems. *Journal of Marketing Research* 37(3): 363-375.

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.

Aral, S., Brynjolfsson, E., Van Alstyne, M. (2007) Productivity effects of information diffusion in e-mail networks. *International Conference on Information Systems*, Montreal, 1-16.

Atahan, P., Sarkar, S. (2011) Accelerated learning of user profiles. *Management Science* 57(2): 215-239.

Bakos, J.Y. (1997) Reducing buyer search costs: Implications for electronic marketplaces. *Management Science* 43(12): 1676-1692.

Barber, B.M., Odean, T. (2008) All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies* 21(2): 785-818.

Barberis, N., Shleifer, A. (2003) Style investing. *Journal of Financial Economics* 68(2): 161-199.

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

Bass, F.M., Wittink, D.R. (1975) Pooling issues and methods in regression analysis with examples in marketing research. *Journal of Marketing Research* 12(4): 414-425.

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

Brancheau, J.C., Wetherbe, J.C. (1990) The adoption of spreadsheet software: Testing innovation diffusion theory in the context of end-user computing. *Information Systems Research* 1(2): 115-143.

Brynjolfsson, E., Smith, M.D. (2000) Frictionless commerce? A comparison of Internet and conventional retailers. *Management Science* 46(4): 563-585.

Calin, M., Palme, E., Sutanto, J., Dellarocas, C. (2012) Attention allocation in information-rich environments: The case of news aggregators. Working paper, Swiss Federal Institute of Technology & Boston University, 1-27.

Cameron, A.C., Gelbach, J.B., Miller, D.L. (2011) Robust inference with multiway clustering. *Journal of Business & Economic Statistics* 29(2): 238-249.

Campbell, N.C.G., Graham, J.L., Jolibert, A., Meissner, H.G. (1988) Marketing negotiations in France, Germany, the United Kingdom, and the United States. *Journal of Marketing* 52(2): 49-62.

Chen, L., Da, Z., Zhao, X. (2013) What drives stock price movements? *Review of Financial Studies* 26(4): 841-876.

Chevalier, J.A., Mayzlin, D. (2006) The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research* 43(3): 345-354.

Choi, H., Varian, H.A.L. (2012) Predicting the present with Google trends. *Economic Record* 88: 2-9.

Cohen, L., Frazzini, A. (2008) Economic links and predictable returns. *The Journal of Finance* 63(4): 1977-2011.

Da, Z.H.I., Engelberg, J., Gao, P. (2011) In search of attention. *The Journal of Finance* 66(5): 1461-1499.

De Nooy, W., Mrvar, A., Batagelj, V. (2005) *Exploratory social network analysis with Pajek* (Cambridge University Press, New York).

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.

Dellarocas, C., Zhang, X., Awad, N.F. (2007) Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing* 21(4): 23-45.

Dellavigna, S., Pollet, J.M. (2009) Investor inattention and friday earnings announcements. *The Journal of Finance* 64(2): 709-749.

Dewan, S., Ramaprasad, J. (2012) Music blogging, online sampling, and the long tail. *Information Systems Research* 23(3-Part-2): 1056-1067.

Dhar, V., Chang, E.A. (2009) Does chatter matter? The impact of user-generated content on music sales. *Journal of Interactive Marketing* 23(4): 300-307.

Dhar, V., Geva, T., Oestreicher-Singer, G., Sundararajan, A. (Forthcoming) Prediction in economic networks. *Information Systems Research*: 1-21.

Diether, K.B., Lee, K.-H., Werner, I.M. (2009) It's SHO time! Short-sale price tests and market quality. *The Journal of Finance* 64(1): 37-73.

Droge, C., Stanko, M.A., Pollitte, W.A. (2010) Lead users and early adopters on the web: The role of new technology product blogs. *Journal of Product Innovation Management* 27(1): 66-82.

Duan, W., Gu, B., Whinston, A.B. (2008a) Do online reviews matter? — an empirical investigation of panel data. *Decision Support Systems* 45(4): 1007-1016.

Duan, W., Gu, B., Whinston, A.B. (2008b) The dynamics of online word-of-mouth and product sales—an empirical investigation of the movie industry. *Journal of Retailing* 84(2): 233-242.

Edgerton, J. (2012) Agency problems in public firms: Evidence from corporate jets in leveraged buyouts. *The Journal of Finance* 67(6): 2187-2213.

Evans, R.B., Geczy, C.C., Musto, D.K., Reed, A.V. (2009) Failure is an option: Impediments to short selling and options prices. *Review of Financial Studies* 22(5): 1955-1980.

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

Froot, K., Teo, M. (2008) Style investing and institutional investors. *Journal of Financial & Quantitative Analysis* 43(4): 883-906.

Ghose, A., Han, S.P. (2011) An empirical analysis of user content generation and usage behavior on the mobile Internet. *Management Science* 57(9): 1671-1691.

Godes, D., Mayzlin, D. (2004) Using online conversations to study word-of-mouth communication. *Marketing Science* 23(4): 545-560.

Graham, J.R., Kumar, A. (2006) Do dividend clienteles exist? Evidence on dividend preferences of retail investors. *The Journal of Finance* 61(3): 1305-1336.

Granados, N., Gupta, A., Kauffman, R.J. (2012) Online and offline demand and price elasticities: Evidence from the air travel industry. *Information Systems Research* 23(1): 164-181.

Green, T.C., Hwang, B.-H. (2009) Price-based return comovement. *Journal of Financial Economics* 93(1): 37-50.

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

Greenwood, R. (2008) Excess comovement of stock returns: Evidence from cross-sectional variation in Nikkei 225 weights. *Review of Financial Studies* 21(3): 1153-1186.

Gu, B., Park, J., Konana, P. (2012) The impact of external word-of-mouth sources on retailer sales of high-involvement products. *Information Systems Research* 23(1): 182-196.

Haugen, R.A., Baker, N.L. (1996) Commonality in the determinants of expected stock returns. *Journal of Financial Economics* 41(3): 401-439.

Hirshleifer, D., Hou, K., Teoh, S.H., Zhang, Y. (2004) Do investors overvalue firms with bloated balance sheets? *Journal of Accounting and Economics* 38(0): 297-331.

Hirshleifer, D., Lim, S.S., Teoh, S.H. (2009) Driven to distraction: Extraneous events and underreaction to earnings news. *The Journal of Finance* 64(5): 2289-2325.

Hirshleifer, D., Teoh, S.H. (2003) Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics* 36(1–3): 337-386.

Hong, H., Lim, T., Stein, J.C. (2000) Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies. *The Journal of Finance* 55(1): 265-295.

Hou, K. (2007) Industry information diffusion and the lead-lag effect in stock returns. *Review of Financial Studies* 20(4): 1113-1138.

Hou, K., Peng, L., Xiong, W. (2008) A tale of two anomalies: The implications of investor attention for price and earnings momentum. Working paper, Ohio State University and Princeton University, 1-43.

Huang, Z., Zeng, D.D., Chen, H. (2007) Analyzing consumer-product graphs: Empirical findings and applications in recommender systems. *Management Science* 53(7): 1146-1164.

Huberman, G., Dorn, D. (2009) Preferred risk habitat of individual investors. Working paper, Columbia Business School and Drexel University, 1-59.

Huberman, G., Regev, T. (2001) Contagious speculation and a cure for cancer: A nonevent that made stock prices soar. *The Journal of Finance* 56(1): 387-396.

James, W. (1981) *The principles of psychology* (Harvard University Press, Cambridge, Mass.).

Jegadeesh, N., Titman, S. (1993) Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance* 48(1): 65-91.

Johnson, E.J., Moe, W.W., Fader, P.S., Bellman, S., Lohse, G.L. (2004) On the depth and dynamics of online search behavior. *Management Science* 50(3): 299-308.

Kahneman, D. (1973) *Attention and effort* (Prentice-Hall, Englewood Cliffs, N.J).

Kalish, S. (1985) A new product adoption model with price, advertising, and uncertainty. *Management Science* 31(12): 1569-1585.

Katona, Z., Zubcsek, P.P., Sarvary, M. (2011) Network effects and personal influences: The diffusion of an online social network. *Journal of Marketing Research (JMR)* 48(3): 425-443.

Kim, J.B., Albuquerque, P., Bronnenberg, B.J. (2011) Mapping online consumer search. *Journal of Marketing Research* 48(1): 13-27.

Kumar, A. (2009a) Dynamic style preferences of individual investors and stock returns. *Journal of Financial and Quantitative Analysis* 44(03): 607-640.

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

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.

Kumar, A., Page, J.K., Spalt, O.G. (2013) Investor sentiment and return comovements: Evidence from stock splits and headquarters changes. *Review of Finance* 17(3): 921-953.

Kuruzovich, J., Viswanathan, S., Agarwal, R. (2010) Seller search and market outcomes in online auctions. *Management Science* 56(10): 1702-1717.

Kuruzovich, J., Viswanathan, S., Agarwal, R., Gosain, S., Weitzman, S. (2008) Marketspace or marketplace? Online information search and channel outcomes in auto retailing. *Information Systems Research* 19(2): 182-201.

Lam, M. (2004) Neural network techniques for financial performance prediction: Integrating fundamental and technical analysis. *Decision Support Systems* 37(4): 567-581.

Leung, A., Agarwal, A., Konana, P., Kumar, A. (2012) Online search: Identifying new investment habitats. *Proceedings of International Conference on Information Systems 2012*, Orlando, Florida, 1-16.

Li, Q., Maggitti, P., Smith, K., Tesluk, P., Katila, R. (2013) Top management attention to innovation: The role of search selection and intensity in new product introductions. *Academy of Management Journal* 56(3): 893-916.

Lin, M., Prabhala, N.R., Viswanathan, S. (2013) Judging borrowers by the company they keep: Friendship networks and information asymmetry in online peer-to-peer lending. *Management Science* 59(1): 17-35.

Luce, R., Perry, A. (1949) A method of matrix analysis of group structure. *Psychometrika* 14(2): 95-116.

Luo, X., Zhang, J., Duan, W. (2013) Social media and firm equity value. *Information Systems Research* 24(1): 146-163.

Manchanda, P., Ansari, A., Gupta, S. (1999) The "shopping basket": A model for multicategory purchase incidence decisions. *Marketing Science* 18(2): 95-114.

Massa, M., Zhang, L. (2009) Cosmetic mergers: The effect of style investing on the market for corporate control. *Journal of Financial Economics* 93(3): 400-427.

Mayzlin, D., Yoganarasimhan, H. (2012) Link to success: How blogs build an audience by promoting rivals. *Management Science* 58(9): 1651-1668.

Mech, T.S. (1993) Portfolio return autocorrelation. *Journal of Financial Economics* 34(3): 307-344.

Mehta, N. (2007) Investigating consumers' purchase incidence and brand choice decisions across multiple product categories: A theoretical and empirical analysis. *Marketing Science* 26(2): 196-217.

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

Moe, W.W., Fader, P.S. (2004) Dynamic conversion behavior at e-commerce sites. *Management Science* 50(3): 326-335.

Moskowitz, T.J., Grinblatt, M. (1999) Do industries explain momentum? *The Journal of Finance* 54(4): 1249-1290.

Niraj, R., Padmanabhan, V., Seetharaman, P.B. (2008) A cross-category model of households' incidence and quantity decisions. *Marketing Science* 27(2): 225-235.

Oestreicher-Singer, G., Sundararajan, A. (2012a) Recommendation networks and the long tail of electronic commerce. *MIS Quarterly* 36(1): 65-A64.

Oestreicher-Singer, G., Sundararajan, A. (2012b) The visible hand? Demand effects of recommendation networks in electronic markets. *Management Science* 58(11): 1963-1981.

Pandit, S., Wasley, C.E., Zach, T. (2011) Information externalities along the supply chain: The economic determinants of suppliers' stock price reaction to their customers' earnings announcements. *Contemporary Accounting Research* 28(4): 1304-1343.

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.

Pindyck, R.S., Rotemberg, J.J. (1993) The comovement of stock prices. *The Quarterly Journal of Economics* 108(4): 1073-1104.

Pirinsky, C., Wang, Q. (2006) Does corporate headquarters location matter for stock returns? *The Journal of Finance* 61(4): 1991-2015.

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

Preis, T., Moat, H.S., Stanley, H.E. (2013) Quantifying trading behavior in financial markets using Google trends. *Scientific Reports* 3(1684).

Rashes, M.S. (2001) Massively confused investors making conspicuously ignorant choices (MCI–MCIC). *The Journal of Finance* 56(5): 1911-1927.

Rosenthal, L., Young, C. (1990) The seemingly anomalous price behavior of royal dutch/shell and unilever n.V./plc. *Journal of Financial Economics* 26(1): 123-141.

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.

Sundararajan, A., Provost, F., Oestreicher-Singer, G., Aral, S. (2013) Information in digital, economic and social networks. *Information Systems Research* 24(4): 883-905.

Susarla, A., Oh, J.-H., Tan, Y. (2012) Social networks and the diffusion of user-generated content: Evidence from YouTube. *Information Systems Research* 23(1): 23-41.

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.

Tetlock, P.C. (2007) Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance* 62(3): 1139-1168.

Thompson, S.B. (2011) Simple formulas for standard errors that cluster by both firm and time. *Journal of Financial Economics* 99(1): 1-10.

Tsai, C.-F., Hsiao, Y.-C. (2010) Combining multiple feature selection methods for stock prediction: Union, intersection, and multi-intersection approaches. *Decision Support Systems* 50(1): 258-269.

Van Roy, B., Yan, X. (2010) Manipulation robustness of collaborative filtering. *Management Science* 56(11): 1911-1929.

Vijh, A. (1994) S&P 500 trading strategies and stock betas. *Review of Financial Studies* 7(1): 215-251.

Wahal, S., Yavuz, M.D. (2013) Style investing, comovement and return predictability. *Journal of Financial Economics* 107(1): 136-154.

Wallace, T.D. (1972) Weaker criteria and tests for linear restrictions in regression. *Econometrica* 40(4): 689-698.

Watkins, J.H. (2007) Prediction markets as an aggregation mechanism for collective intelligence. *UCLA Lake Arrowhead Conference on Human Complex Systems*, Los Angeles, CA, 1-9.

Weber, T.A., Zheng, Z. (2007) A model of search intermediaries and paid referrals. *Information Systems Research* 18(4): 414-436.

Wu, L., Brynjolfsson, E. (2009) The future of prediction: How Google searches foreshadow housing prices and sales. *Proceedings of Thirtieth International Conference on Informaiton Systems*, Phoenix, 1-14.

Zeng, X., Wei, L. (2013) Social ties and user content generation: Evidence from Flickr. *Information Systems Research* 24(1): 71-87.

Zhu, F., Zhang, X. (2010) Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of Marketing* 74(2): 133-148.

Vita

Alvin Leung received his Bachelor of Business Administration (Information Systems), Bachelor of Engineering (Software Engineering), and Master of Philosophy from the University of Hong Kong in 2005, 2006, and 2009, respectively. He joined the PhD program of Department of Information, Risk, and Operations Management, McCombs School of Business in August 2009. In May 2011, he obtained Master of Science in Information, Risk, and Operations Management. Alvin has published several papers in the area of information security and online communities in information systems and computer science journals such as Decision Support Systems, the Communications of the ACM, Computers in Human Behavior and the Communications of the AIS. He has also presented his research in leading information systems conferences such as International Conference on Information Systems and Workshop on Information Systems and Economics.

Permanent address (or email): 2110 Speedway Stop B6500 Austin TX 78712
aleung@utexas.edu

This dissertation was typed by the author.