

A Study of Search Attention and Stock Returns Cross Predictability

Full paper

Alvin Chung Man Leung
City University of Hong Kong
acmleung@cityu.edu.hk

Prabhudev Konana
The University of Texas at Austin
prabhudev.konana@mcombs.utexas.edu

Ashish Agarwal
The University of Texas at Austin
ashish.agarwal@mcombs.utexas.edu

Alok Kumar
University of Miami
akumar@bus.miami.edu

Abstract

This study investigates a novel application of correlated online searches in predicting stock performance across supply chain partners. If two firms are economically dependent through supply-chain relationship and if information related to both firms diffuses in the market slowly (rapidly), then our ability to predict stock returns increases (vanishes). Using supply-chain data and weekly co-search network of supply-chain partners from Bloomberg and Yahoo! Finance, respectively, we find that when investors of a focal stock pay less attention to its supply-chain partners, we can use lagged partner returns to predict the future return of the focal stock. When investors' co-attention to focal and partner stocks is high, the predictability is low. We contribute to the growing literature on aggregate search and economics of networks by demonstrating the inferential power and economic implications of search networks.

Keywords

Attention, search, stock, supply-chain, prediction.

Introduction

Online search activity is used as a proxy to measure the level of interest or attention about a product or an asset. The search volume or trend is then used for predicting demand (Choi and Varian 2012), house prices (Wu and Brynjolfsson 2009) and stock price returns (Da et al. 2011; Luo et al. 2013). The ability to predict using massive amounts of search results has attracted significant interest among researchers and practitioners. This research investigates a novel approach to correlated searches – that is, search related to multiple items – in predicting stock returns among firms with supply chain relationships.

The basic idea is that stocks of economically linked firms such as supply chain partners are strongly correlated due to correlated fundamentals (Hong et al. 2007) and profits (Menzly and Ozbas 2010). Thus, investors should be paying equal attention to supply chain partners, which may be reflected in online correlated searches in a given period. However, as a result of limited attention and investor specialization, it is possible that the information diffuses slowly in the market (Hong et al. 2007) and even across economically linked assets (Cohen and Frazzini 2008; Menzly and Ozbas 2010). As a consequence, while the supply chain partner stocks are correlated, there may be lag in this correlation and the lagged returns of partner firms can be used to predict the current returns of a focal firm (Cohen and Frazzini 2008; Hou 2007; Menzly and Ozbas 2010). The main thesis of this paper is that if there is variation in information diffusion among supply chain partners, online correlated searches of partner firms should also reflect

such a behavior. That is, online correlated search is a proxy for investor attention of supply chain partners. If this is true then online correlated searches can be used to predict stock returns.

If users pay attention to the stocks of supply chain partners in the same time period, it represents higher level of information diffusion that would lead to return comovement for such stocks in the same time period (Barberis et al. 2005). In that case, lagged correlation in stocks is less likely among partner stocks with high level of co-attention as compared to lagged correlation among partner stocks with low level of co-attention. The low level of co-attention for some partners may occur due to limited cognitive resources to process and search extensive number of assets (Kahneman 1973; Li et al. 2013). Thus, correlated searches for stocks can reveal the extent of investor co-attention and can be proxy for the extent of information diffusion. Further, as users' attention change, the nature of correlated search patterns may also evolve, which reflect that over time there may be varying intensity of information diffusion across supply chain partners.

When investors search for various stocks, IT platforms capture digital footprints of correlated searches (i.e., search history). The correlated search data can be used to construct a network of stocks that users search akin to product network where associations or links between assets are formed due to joint economic activity for the assets (Dhar et al. 2014; Oestreicher-Singer and Sundararajan 2012b). Such network information can be used to make market-level predictions (Dhar et al. 2014) for individual assets. Similarly, a search network based on correlated searches can reveal level of co-attention and information diffusion across supply chain partners and can potentially be used for predicting the returns of individual stocks in the supply chain.

There could be several unknown factors that drive these user searches resulting in the search network. These unknown factors may vary across different supply chain networks. An alternative approach would be to explicitly investigate all possible unknown factors and use these as a proxy for investor attention to determine cross-predictability of supply chain stocks. Dhar et al. (2014) argue that a network representation may constitute a decision-relevant "projection" of complex space of unknown factors into a pattern of correlated observable activities. Search network for complementary assets represents a similar network. Sundararajan et al. (2013) argue for the need to understand and evaluate the economic value of such networks revealed through IT. One could argue that it is easy to deduce co-attention or inattention among supply chain partners through actual transactions of trades. However, transaction information is private and not available publicly. Correlated searches can be extracted from public IT portals and can reveal the extent of information diffusion across supply chain partners' stocks, which can then be used for cross-prediction.

Using the online co-search data from Yahoo! Finance of Russell 3000 index stocks, we construct a correlated search network for the supply chain stocks, where the nodes represent stocks and the edges represent the co-search intensity across stocks of supply chain partners. We determine the supply chain partners from Bloomberg Supply Chain Analysis (SPLC) module. 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 control for the effects of commonly used risk factors and news related to the stocks that can impact the returns. We also control for factors that can drive investor attention such as institutional holdings and analyst coverage, and factors that influence co-search for supply chain partners such as news co-mentions and investment styles. We also verify that our results hold even after accounting for unobservable cross-sectional differences across partners which could potentially drive the outcome.

Our results suggest that the online co-search intensity across supply chain stocks could be a proxy for the extent of information diffusion. High co-search intensity may represent high information diffusion and as a consequence results in weaker cross predictability of stocks. Further, our results show that this effect persists even after accounting for the known drivers of such information diffusion.

Our study makes several contributions. From a managerial perspective, it showcases how prediction capabilities can be improved with consumer co-search data. As online activities now become more prevalent and transparent, firms can access co-searching data (which are available in cookies, server log files and search queries) more easily and analyzed in real-time rather than post-hoc evaluation.

We add to the IS and the finance literature on using aggregate search for an asset to predict its market performance (Choi and Varian 2012; Da et al. 2011; Luo et al. 2013; Preis et al. 2013; Wu and Brynjolfsson 2009). Our study shows how correlated search across economically linked assets such as supply chain stocks can be used to predict the market performance of individual assets. We also add to the emerging literature on economic networks which has considered correlated purchases to identify aggregate user preferences for products (Oestreicher-Singer and Sundararajan 2012a; Oestreicher-Singer and Sundararajan 2012b; Oestreicher-Singer and Zalmanson 2013) and to make predictions (Dhar et al. 2014). We demonstrate how such networks can be used to qualify existing economic associations between assets or products as they reveal active user attention and can help improve the cross predictability across such assets.

In addition, our study demonstrates the value of slow information diffusion in a network and the role of co-searches in identifying the level of information diffusion. Most network studies focus on the fast information diffusion and identifying the network properties to facilitate faster information diffusion. Our study shows that in a network of economically linked assets, slow information diffusion paths can be identified using co-search intensity and can be utilized to improve predictions.

Finally, our study also contributes to finance research on cross predictability of stocks. Most prior studies in this area (e.g. Cohen and Frazzini 2008) rely on passive measures of investor co-attention such as institutional ownership and analyst coverage to show extent of information diffusion across supply chain stocks. Our study shows how co-search intensity across supply chain partners is a more active measure of investor co-attention to represent the level of information diffusion across supply chain stocks. Further, our study shows that this measure can be used to evaluate the cross predictability at a more granular level instead of the industry level analysis in the existing finance literature (Cohen and Frazzini 2008; Menzly and Ozbas 2010).

Methodology

In this study, we build a prediction model based on the variation of co-search intensity and supply-chain strength across economically linked stocks. We explain each of them in the following subsections.

Co-search Intensity

We use Yahoo! “also-viewed” data to capture the co-search pattern across supply chain stocks. Yahoo! Finance 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.¹ It has an average monthly traffic of over 45 million visitors². Additionally, among the investment portals with a similar scale of visitors, Yahoo! Finance is the only portal that reveals the co-viewing pattern of investors.³ Yahoo! Finance lists top six co-viewed stocks for each stock on the stock summary page. Figure 1 shows an example of Yahoo! Finance stock summary page. The circled area shows the top six “also-viewed” stocks⁴. When the majority of Yahoo! users who search stock A (e.g. AMD in Figure 1) also search B (e.g. INTC in Figure 1), stock B is in the “also-viewed” list of stock A. These co-viewed stocks may also include supply chain partners. For example, DELL and AMD are supply chain partners. Yahoo! computes this co-viewed data based on visitors’ cookies⁵ and uses a threshold to upload the most recent data to Yahoo! Finance.⁶ Using a Perl

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

² <http://www.ebizmba.com/articles/business-websites>

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

⁴ 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.

⁵ Cookies allow a website to identify and track all user activities, including search for different items (in our case stocks).

script, we collected daily co-viewing data for all Russell 3000 stocks at 4pm CST every day during the period from September 15, 2011 to December 31, 2013.

We use the co-viewing data to identify subsets of partner stocks that attract investor attention during a certain time period. 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 by the end of year 2011 because those thinly traded stocks are more volatile to market changes and may confound our cross-predictability results (Menzly and Ozbas 2010). Our estimation sample contains 102,910 firm-week data that comprise of 1,619 focal firms in 66 trading weeks for the period from September 15, 2011 to December 31, 2012.

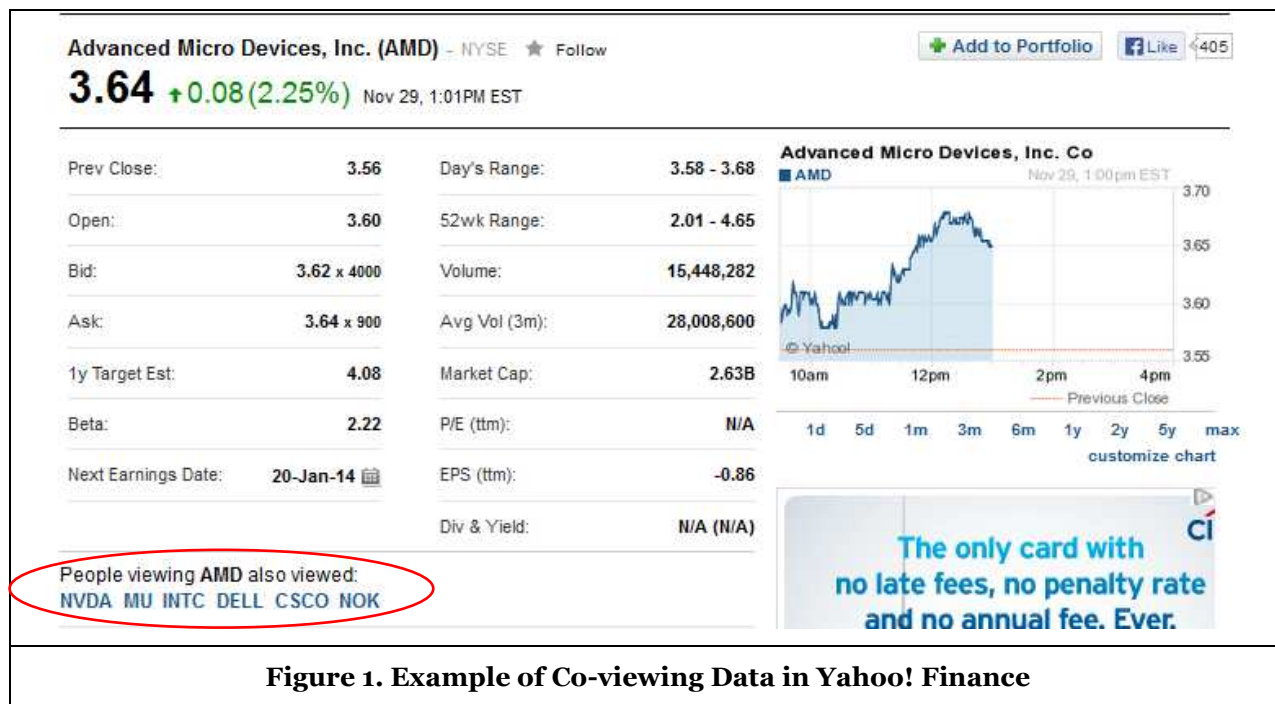


Figure 1. Example of Co-viewing Data in Yahoo! Finance

Based on the co-viewing pattern, we can construct a directed graph as shown in Figure 2. We use the example of Advanced Micro Devices (AMD) in Figure 1 as an example. The arrow represents potential information flow from a frequently co-searched stock to a focal stock. In our example, AMD has six “also-viewed” stocks on November 29, 2013. It means that investors who search AMD may also receive information of six other stocks.

Supply-chain Strength

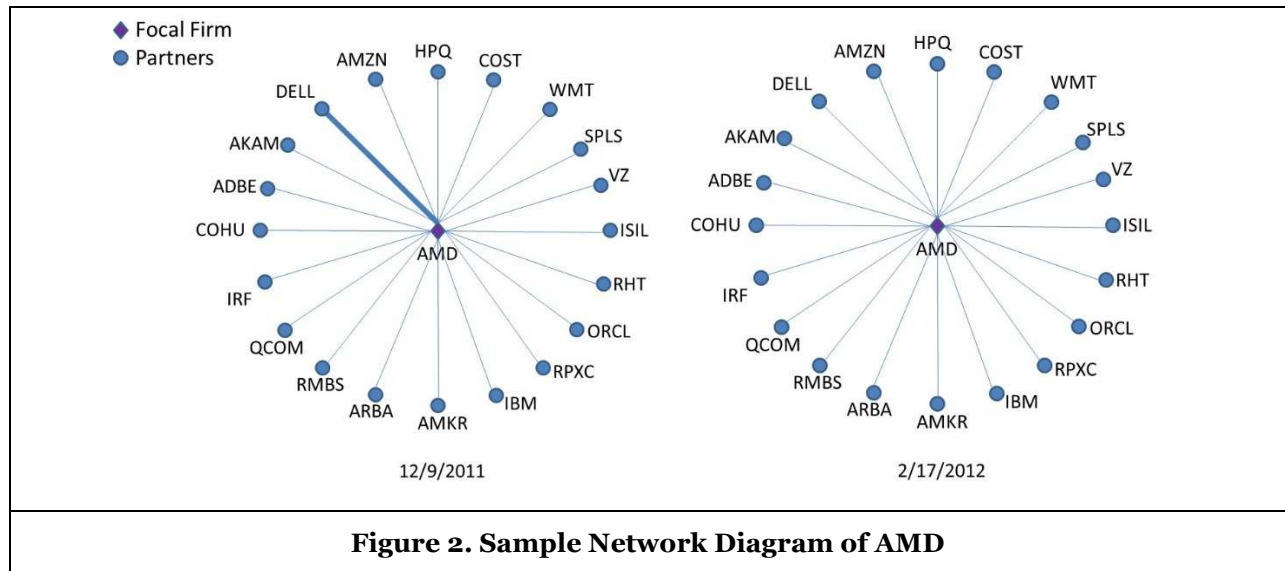
Prior studies (e.g, Cohen and Frazzini 2008; Menzly and Ozbas 2010) have shown that supply-chain relationship may influence cross-predictability. To control for the potential differential impact of different types of supply-chain partners, we split them into suppliers and buyers. We identify supplier and buyer relationships among Russell 3000 stocks based on Bloomberg Supply Chain Analysis (SPLC). Using SPLC, we identify all supplier and buyer pairs and retrieve trading amount between each pair. The trading data are based on industrial estimates by Bloomberg analysts as well as the data reported by firms in the quarterly earnings. SPLC also provides data of revenue percentage and cost percentage, which are similar to Pandit et al. (2011)’s definition of supplier dependency and customer exposure. Supplier dependency is defined as the trading amount between a supplier and a buyer divided by the total revenue of the supplier;

⁶ 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.

customer exposure is defined as the trading amount between the two parties divided by the total cost of goods sold to the customer. To control the potential influence of supply-chain intensity on the user attention, we compute supply-chain strength weighted lagged partner returns and analyze whether such returns can help in predicting contemporary returns of focal firms.

Research Data

We focus on Russell 3000 stocks listed from September 2011 to December 2012 with valid supply-chain data from SPLC. Russell 3000 stocks account for 98% of the total market capitalization of all stocks trading in the U.S. They are commonly used in the recent literature on stock returns (Da et al. 2011; Diether et al. 2009; Haugen and Baker 1996). Furthermore, we restrict our data to large firms with market capitalization above 20th percentile of stocks listed in NYSE by the end of year 2011. The primary reason is that small firms are more responsive to news and they may bias the results of cross-predictions. Such filtering can eliminate thin market effect and has been adopted in prior cross-prediction studies (e.g., Menzly and Ozbas 2010). Our final sample data consist of 1,620 Russell 3000 stocks with valid stock return and supply-chain strength data. Figure 2 shows a typical example of co-search relationship between AMD and its core partners on two different dates. A thick edge implies the majority of investors who search AMD also search its partner in a particular week (e.g. a week ending on 12/9/2011).



Research Model

We analyze the cross-predictability of supply-chain partners using the approach adopted by Menzly and Ozbas (2010). Specifically, we estimate the following time series model:

$$\begin{aligned}
 Ret_{i,t} = & \beta_0 + \beta_1 Ret_{p,t-1}^H + \beta_2 Ret_{p,t-1}^L + \beta_3 Ret_{i,t-1} + \beta_4 MktRf_t + \beta_5 SMB_t + \beta_6 HML_t + \beta_7 MOM_t + \\
 & \beta_8 Analyst_{i,t-1} + \beta_9 InstHldg_{i,t-1} + \beta_{10} News_{i,t} + \beta_{11} News_{i,t-1} + \beta_{12} CoNews_{p,t}^H + \beta_{13} CoNews_{p,t-1}^H + \\
 & \beta_{14} CoNews_{p,t}^L + \beta_{15} CoNews_{p,t-1}^L + \varepsilon_{i,t}
 \end{aligned}
 \tag{1}$$

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). $Ret_{p,t-1}^H$ and $Ret_{p,t-1}^L$ are supply-chain strength weighted partner returns with 1 week lag for the high and low co-search intensity partners. If the partner returns can predict the returns of the focal stock then we should expect the coefficient of $Ret_{p,t-1}^H$ to be positive and significant.

If supply-chain partners of a focal firm are listed in the co-searching list of the focal firm at least one day of the previous week, we consider these as part of the high co-search intensity group (H). Otherwise, they

are categorized as part of the low co-search intensity group (L). We compute supply-chain strength (SC) weighted average partner returns separately for both groups. 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) use the same approach to compute composite partner returns.

If a focal firm has both buyers and suppliers as its supply-chain partners, we define $R_{P,t-1}$ as the average of dependency weighted buyer returns $R_{B,t-1}$ (Equation 2) and exposure weighted customer returns $R_{S,t-1}$ (Equation 3). Dependency is the trading amount between a focal firm and a buyer divided by the total revenue of the focal firm and exposure is the trading amount between the focal firm and a customer divided by the total cost of goods sold of the focal firm. If the focal firm only has one type of partners, then we use the corresponding supply chain (SC) weighted average partner return (i.e. $R_{B,t-1}$ or $R_{S,t-1}$) as $R_{P,t-1}$.

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

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

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 obtained the stock return data for each stock during the study period from the Center for Research on Security Prices (CRSP) database.

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 using the Fama-French 4 factors (i.e., *MktRf*, *SMB*, *HML* and *MOM*). Prior studies (e.g., Menzly and Ozbas 2010) show that analyst coverage (*Analyst*) and institutional ownership (*InstHldg*) also represent attention and can influence a stock's return. We also control for those factors. We measure *Analyst* as $\log(1 + \text{the number of analysts following the focal stock in period } t)$. We obtain the analyst coverage information from the I/B/E/S database. We measure *InstHldg* as $\log(1 + \text{percentage of institutional holding in period } t)$.⁷ We obtain the institutional holding data 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 the current week $News_{i,t}$ and news in previous week, $News_{i,t-1}$. We calculate news volume as $\log(1 + \text{total number of news articles related to the focal firms})$. Da et al. (2011) uses similar approach in their calculation of news volume.

In addition to the news for the focal stock, co-mentions of the focal stock with its supply chain partners in the news can also influencing the returns for the focal stock. For example, it may be the case that investors co-search a focal stock with some of its supply chain partners because of such news co-mentions. In that case, one can just use the news co-mentions instead of co-searches to explain the cross-predictability across supply chain partners. In order to verify if that is the case, we also control for the effect of news co-mentions across supply chain partners. We control for news co-mentions for each co-search intensity group in the current week (i.e., $CoNews_{P,t}^L$ and $CoNews_{P,t}^H$) and the previous week (i.e., $CoNews_{P,t-1}^L$ and $CoNews_{P,t-1}^H$).

We obtain news volume and co-mention news volume from all sources of news available in Factiva news database. We count the number of articles in the Factiva database associated with co-mentions of focal stock and each partner.⁸ We weigh the number of co-mentions for each pair with the supply chain

⁷ Data for number of analysts and the percentage of institutional holding is available on a quarterly basis. We assume that it does not change within a quarter and is applicable for all weeks within a quarter.

⁸ We search individual stock on Factiva in each time period and download the list of top most mentioned companies on Factiva. The top most mentioned companies are companies that are also mentioned with the focal stock appearing in the news database. Apart from the name of companies that are co-mentioned with the focal stock, the number of

strength. $CoNews_{P,t}$ is $\log [1 + (CoNews_{BP,t} + CoNews_{SP,t})/2]$. If a focal firm has only one type of partners, then we either use $\log (1 + CoNews_{BP,t})$ or $\log (1 + CoNews_{SP,t})$, where $CoNews_{BP,t} = \sum_{j \in B} Dep_{ij} \times News_{i,j,t} / \sum_{j \in B} Dep_{ij}$ and $CoNews_{SP,t} = \sum_{j \in S} Exp_{ij} \times News_{i,j,t} / \sum_{j \in S} Exp_{ij}$. $News_{i,j,t}$ is the total number of news articles with both companies i and j at time t .

Results

Table 1 shows summary statistics of our sample data and Table 2 shows correlation matrix. The correlation among independent variables is low except for the news variables. However, the variation inflation factor (VIF) of our regression result is less than 6 suggesting that multi-collinearity is not an issue. Nevertheless, we have also tried combining contemporary news and lagged news together and re-run the regression. The VIF is below four and the research findings are qualitatively similar. We estimate our research model using two-dimensional clustering at firm and week level. Two dimensional clustering is a commonly used approach in finance to account for cross-sectional correlation and auto-correlation in the analysis of stock returns (Petersen 2009).

Variable	N	Mean	SD	Min	Max
$Ret_{P,t-1}^I$	102,910	0.00	0.06	-0.74	1.49
$Ret_{P,t-1}^H$	102,910	0.00	0.04	-0.66	0.92
$Ret_{i,t-1}$	102,910	0.00	0.02	-0.48	0.47
$MktRf_t$	102,910	0.00	0.06	-0.74	1.49
SMB_t	102,910	0.00	0.02	-0.05	0.08
HML_t	102,910	0.00	0.01	-0.03	0.03
$ MOM_t$	102,910	0.00	0.01	-0.02	0.02
$Analyst_{i,t-1}$	102,910	0.00	0.02	-0.04	0.03
$InstHldg_{i,t-1}$	102,910	1.66	0.78	0.00	3.91
$News_{i,t}$	102,910	0.56	0.13	0.00	1.20
$News_{i,t-1}$	102,910	2.44	1.39	0.00	8.64
$CoNews_{P,t}^I$	102,910	2.39	1.46	0.00	8.64
$CoNews_{P,t-1}^I$	102,910	0.11	0.34	0.00	5.42
$CoNews_{P,t}^H$	102,910	0.07	0.30	0.00	5.42
$CoNews_{P,t-1}^H$	102,910	0.13	0.53	0.00	6.06
$Ret_{P,t-1}^I$	102,910	0.11	0.51	0.00	6.06

Table 1. Summary Statistics

news articles that mention both companies are also shown. We use this information for the calculation of news co-mention volume.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
(1) $Ret_{i,t}$	1	-0.03	-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) $Ret_{p,t-1}^I$	-0.02	1	0.22	0.46	-0.08	0.09	-0.13	0.00	-0.01	0.00	0.01	0.01	0.01	0.01	0.00	0.00
(3) $Ret_{p,t-1}^H$	0.00	0.21	1	0.21	-0.03	0.06	-0.06	0.01	0.00	0.00	0.02	0.02	0.02	0.02	0.05	0.04
(4) $Ret_{i,t-1}$	-0.04	0.38	0.20	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	-0.07	-0.02	-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	0.07	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	-0.09	-0.05	-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	0.02	0.02	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	-0.01	0.00	-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	0.01	-0.02	0.00	-0.01	0.02	0.26	-0.04	0.82	1	0.40	0.40	0.29	0.30
(13) $CoNews_{p,t}^I$	0.01	0.01	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) $CoNews_{p,t-1}^I$	0.00	0.00	0.01	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) $CoNews_{p,t}^H$	0.00	0.00	0.03	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) $CoNews_{p,t-1}^H$	0.00	0.00	0.03	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 2. Correlation Matrices

Table 3 shows the main results. Column 1 shows the parameter estimates using the two dimensional clustering. The coefficient of lagged partner returns of low co-search intensity group is significant and positive. However, the coefficient of lagged partner returns of high co-search intensity group is insignificant. These results suggest that there is a lagged reaction to the partner stocks where the co-search intensity is low. However, there is no lagged reaction to the partner stocks with high co-search intensity. Lagged reaction to partner stocks is expected due to the slow information diffusion across supply chain stocks (Cohen and Frazzini 2008; Menzly and Ozbas 2010). However, we observe this only for the partner stocks where the co-search intensity is low. A plausible explanation is that the information diffusion is high across partners with high co-search intensity. Thus, co-search intensity can reveal the extent of information diffusion and can be used to determine the cross-predictability among supply chain stocks.

The lagged return of focal stocks is significant and negative. This represent the short-term reversion as documented in earlier research (Jegadeesh and Titman 1993). The Fama-French four factors are all significant except HML. Further, $Analyst_{i,t-1}$ and $InstHdg_{i,t-1}$ are insignificant implying that the two measures for attention have limited impact on weekly predictions.

The coefficient for current focal stock news is significant and positive. It shows that investors of focal stocks are aware of the news of stocks they invest and take immediate action. The coefficient of lagged focal news is significant and negative. This is similar to the effect of lagged focal stock return due to short-term reversion. The current and lagged co-mention news coefficients are insignificant implying that investors, in general, do not incorporate supply-chain partner news in their focal stock valuation.

We repeat the analysis using Fama-MacBeth regression with Newey-West correction for auto-correlation. It is another common method used in Finance literature for times series regression (Petersen 2009) and has been also used for cross predictability analysis (Menzly and Ozbas 2010). In Fama Macbeth regression procedure, we run cross-sectional regressions for different time periods and then use these estimates for individual time periods to derive the overall estimates. Results are shown in column 2 of Table 3 and are qualitatively similar.

Variables	(1)	(2)
$Ret_{P_i,t-1}^L$	0.0174** (0.0070)	0.0392*** (0.0067)
$Ret_{P_i,t-1}^H$	0.0066 (0.0149)	-0.0509 (0.0635)
$Ret_{i,t-1}$	-0.0214** (0.0090)	-0.0479*** (0.0036)
$MktRf_t$	1.0677*** (0.0156)	1.0566*** (0.0147)
SMB_t	0.6194*** (0.0280)	0.6032*** (0.0234)
HML_t	-0.0157 (0.0339)	-0.0375 (0.0275)
MOM_t	-0.1041*** (0.0255)	-0.1242*** (0.0215)
$Analyst_{i,t-1}$	-0.0004	0.0015*

	(0.0004)	(0.0009)
$InstHldg_{i,t-1}$	0.0022 (0.0021)	-0.0303 (0.0716)
$News_{i,t}$	0.0015*** (0.0005)	0.0025*** (0.0005)
$News_{i,t-1}$	-0.0011*** (0.0004)	-0.0001 (0.0003)
$CoNews_{p,i,t}^L$	0.0007 (0.0018)	0.0062 (0.0105)
$CoNews_{p,i,t-1}^L$	-0.0004 (0.0015)	-0.0604 (0.0811)
$CoNews_{p,i,t}^H$	-0.0004 (0.0011)	-0.0002 (0.0011)
$CoNews_{p,i,t-1}^H$	-0.0002 (0.0010)	-0.0101 (0.0081)
Constant	-0.0012 (0.0019)	0.0151 (0.0281)
	Two Dimensional Clustering	Fama-MacBeth
N	102,910	102,910
R ²	0.2545	0.1794

Table 3. Cross-prediction results

Discussion and Conclusion

This study extends extant literature on online user search by focusing on correlated searches across economically linked assets and investigating its usefulness for cross predictability of stocks. We analyze correlated searches for supply chain stocks on Yahoo! Finance data. We find that cross predictability of stocks depends on the co-search intensity. Lagged returns of partners that are not co-searched with the focal stock can be used to predict the returns of a focal stock. However, the same does not hold for partners which are co-searched with the focal stock. Lack of cross predicability across partners suggests faster information diffusion. Thus, our results show that co-search intensity can be a proxy for the extent of information diffusion across supply chain stocks. This is an exciting insight, since past research in finance primarily shows evidence of limited information diffusion across the entire supply chain network but does not distinguish among stocks within the network. Our findings have practical economic value. We show that the past returns of low co-attention supply-chain partners can help predict future returns of a focal stock with control of other known stock return predictors. A prediction model can be developed based on supply-chain partners' co-attention level.

This study has important implications. Our study illustrates the economic value of capturing and analyzing publicly available massive online investor data for investment decisions. Such information can reveal more details about the economic activity and market performance and help make better decisions. More specifically, our study shows that online co-search data such as the Yahoo! Finance “also-viewed” list has several advantages over other measures of co-attention used in prior finance research. Many of the existing measures are either passive or indirect. For example, Cohen and Frazzini (2008) use mutual

funds' joint holdings of supplier/customer stocks as proxies for investor attention. However, mutual funds don't represent the actual attention of retail investors. Other conventional attention proxies include news, extreme past returns and trading volume (Barber and Odean 2008; Hou et al. 2008) are indirect measures of attention. Online investor data such as correlated searches provides a more active measure of investor attention. Further, co-search data is available publicly and can be used to capture user economic activities at a granular level. Traditional publicly available data in finance cannot reveal investor activities at a granular level. In addition, detailed data is typically proprietary and is fragmented across multiple traders.

This study also demonstrates how the variation in the information diffusion across a network of economically linked nodes or assets can be utilized to make predictions. Further, it shows how investor co-search intensity across nodes or assets can be used to detect such variation in the level of information diffusion. The underlying premise is that information flows across nodes due to economic linkages such as supply chain. Thus, if a supply chain partner is not co-searched with a focal stock then the information diffusion is likely to be slow which in turn improves the cross-predictability. In that, our study makes a unique contribution to network analysis, as a typical network analysis focuses on identifying the network linkages to facilitate faster information diffusion. Our study illustrates the value of slow information diffusion. This can be extended to determine cross-predictability across other economically linked assets such as competitors, complementary industries and also complementary markets.

There are several limitations in our analysis that can be the basis for future research. We determine co-search intensity based on the co-appearance of stocks on Yahoo! Finance. However, we do not consider the actual co-search volume, which can help further differentiate between supply chain partners in terms of the co-search intensity. Future research should explore other data sources such as message boards to better measure the co-search intensity across partners and use the precise measure to determine the cross-predictability of stocks. Also, our dataset reveals only the search data, and not the actual transaction data. This analysis can be further improved if access to transactional data is also available for the same user base. In addition, we only demonstrate the predictability of co-attention among aggregated supply-chain partners. Future research may analyze the impact of upstream and downstream partners. When more co-search data become available, we may further analyze the predictability of firms based on supply-chain partners in different time periods. We may also test whether the results hold when we analyze only top 3 co-search supply-chain partners. Last, it would also be useful to identify unknown factors that drive investor co-search pattern and stock cross predictability.

Acknowledgements

The work described in this paper was partially supported by a grant from City University of Hong Kong (Project No. 7200411).

REFERENCES

- Barber, B.M., and 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), pp. 785-818.
- Barberis, N., Shleifer, A., and Wurgler, J. 2005. "Comovement," *Journal of Financial Economics* (75:2), pp. 283-317.
- Choi, H., and Varian, H.R. 2012. "Predicting the Present with Google Trends," *Economic Record* (88), pp. 2-9.
- Cohen, L., and Frazzini, A. 2008. "Economic Links and Predictable Returns," *The Journal of Finance* (63:4), pp. 1977-2011.
- Da, Z.H.I., Engelberg, J., and Gao, P. 2011. "In Search of Attention," *The Journal of Finance* (66:5), pp. 1461-1499.
- Dhar, V., Geva, T., Oestreicher-Singer, G., and Sundararajan, A. 2014. "Prediction in Economic Networks," *Information Systems Research* (25:2), pp. 264-284.
- Diether, K.B., Lee, K.-H., and Werner, I.M. 2009. "It's Sho Time! Short-Sale Price Tests and Market Quality," *The Journal of Finance* (64:1), pp. 37-73.

- Haugen, R.A., and Baker, N.L. 1996. "Commonality in the Determinants of Expected Stock Returns," *Journal of Financial Economics* (41:3), pp. 401-439.
- Hong, H., Torous, W., and Valkanov, R. 2007. "Do Industries Lead Stock Markets?," *Journal of Financial Economics* (83:2), pp. 367-396.
- Hou, K. 2007. "Industry Information Diffusion and the Lead-Lag Effect in Stock Returns," *Review of Financial Studies* (20:4), pp. 1113-1138.
- Hou, K., Peng, L., and Xiong, W. 2008. "A Tale of Two Anomalies: The Implications of Investor Attention for Price and Earnings Momentum." Ohio State University and Princeton University, pp. 1-43.
- Jegadeesh, N., and Titman, S. 1993. "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency," *The Journal of Finance* (48:1), pp. 65-91.
- Kahneman, D. 1973. *Attention and Effort*. Englewood Cliffs, N.J: Prentice-Hall.
- Li, Q., Maggitti, P., Smith, K., Tesluk, P., and 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), pp. 893-916.
- Luo, X., Zhang, J., and Duan, W. 2013. "Social Media and Firm Equity Value," *Information Systems Research* (24:1), pp. 146-163.
- Mech, T.S. 1993. "Portfolio Return Autocorrelation," *Journal of Financial Economics* (34:3), pp. 307-344.
- Menzly, L., and Ozbas, O. 2010. "Market Segmentation and Cross-Predictability of Returns," *The Journal of Finance* (65:4), pp. 1555-1580.
- Oestreicher-Singer, G., and Sundararajan, A. 2012a. "Recommendation Networks and the Long Tail of Electronic Commerce," *MIS Quarterly* (36:1), pp. 65-83.
- Oestreicher-Singer, G., and Sundararajan, A. 2012b. "The Visible Hand? Demand Effects of Recommendation Networks in Electronic Markets," *Management Science* (58:11), pp. 1963-1981.
- Oestreicher-Singer, G., and Zalmanson, L. 2013. "Content or Community? A Digital Business Strategy for Content Providers in the Social Age," *MIS Quarterly* (37:2), pp. 591-616.
- Pandit, S., Wasley, C.E., and 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), pp. 1304-1343.
- Petersen, M.A. 2009. "Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches," *Review of Financial Studies* (22:1), pp. 435-480.
- Preis, T., Moat, H.S., and Stanley, H.E. 2013. "Quantifying Trading Behavior in Financial Markets Using Google Trends," *Scientific Reports* (3:1684).
- Rosenthal, L., and Young, C. 1990. "The Seemingly Anomalous Price Behavior of Royal Dutch/Shell and Unilever N.V./PLC," *Journal of Financial Economics* (26:1), pp. 123-141.
- Sundararajan, A., Provost, F., Oestreicher-Singer, G., and Aral, S. 2013. "Information in Digital, Economic and Social Networks," *Information Systems Research* (24:4), pp. 883-905.
- Wu, L., and Brynjolfsson, E. 2009. "The Future of Prediction: How Google Searches Foreshadow Housing Prices and Sales," *Proceedings of Thirtieth International Conference on Information Systems*, Phoenix, pp. 1-14.