東京大学 工学系研究科 総合研究機構
イノベーション政策研究センター

IPRC Discussion Paper Series No.10

武田 史子
東京大学大学院工学系研究科

若尾 拓実
東京大学大学院工学系研究科

Google search intensity and its relationship with returns and trading volume of Japanese stocks*

September 2013

Innovation Policy Research Center,
Institute of Engineering Innovation, School of Engineering, The University of Tokyo
http://ipr-ctr.t.u-tokyo.ac.jp/

Google search intensity and its relationship with returns and trading volume of Japanese stocks*

Fumiko Takeda+
Department of Technology Management for Innovation, University of Tokyo
E-mail: takeda@tmi.t.u-tokyo.ac.jp

Takumi Wakao
Department of Systems Innovation, University of Tokyo
E-mail: zinc515@gmail.com

ABSTRACT

This paper examines the relationship between online search intensity and stock-trading behavior in the Japanese market. The search intensity is measured by the search volume of company names on Google, which is expected to be related to the aggregate stock purchasing behavior of individual investors. Our sample consists of 189 stocks included in the Nikkei 225 and searched between 2008 and 2011. We find correlations with search intensity that are strongly positive for trading volume and weakly positive for stock returns. Our results are consistent with the notion that the increase of search activity is associated with increases of trading activity, but the probability that this increase of trading raises stock prices is not high, probably because of the fact that our sample period includes major negative economic shocks such as the 2008 world financial crisis and the 2011 Great East Japan Earthquake; also, the presence of individual investors, whose online search activity is expected to be well-associated with stock trading, is smaller in Japan than in the U.S.

JEL classification: G12, G14
Keywords: Investor sentiment; attention: stock price; trading volume; behavioral finance

---

* We would like to thank Yutaka Matsuo for his helpful comments. All remaining errors are of our own.

+ Author in correspondence. Department of Technology Management for Innovation, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656 Japan. Tel/fax: +81-3-5841-1191.

1. Introduction

The purpose of this paper is to examine the relationship between investors' attention and asset prices using Japanese data. Traditionally, many asset pricing models are based on the efficient market hypothesis. In an efficient market, market prices reflect all available information (Fama 1976). However, in the real world, investors do not always have access to all information, but only the limited information they are interested in, as attention is a scarce cognitive activity (Kahneman 1973). This fact may undermine the efficient market hypothesis and invites the question of whether and to what extent market prices reflect investors' interests.

According to Merton (1987), investors' attention is related to the determination of stock prices and liquidity. However, it is difficult to measure the degree of investors' attention. In the present study, we use Google Trends, a public online service based on Google Search, one of the top internet search engines.[1] Google Trends shows how frequently a particular word is searched relative to the total volume of searches in specified periods and locations. Recent empirical finance literature has documented that the search intensity obtained by Google Trends is positively related to stock returns and trading volume (Da et al. 2011; Joseph et al. 2011; Bank et al. 2011; Vlastakis and Markellos 2012), supporting the "price pressure hypothesis" or "attention theory" proposed by Barber and Odean (2008).[2]

The present study complements these prior studies by examining the relationship between search intensity and stock-trading behavior in the Japanese market. We find that search intensity correlates significantly and positively with stock prices and trading volume. However, the significance level is stronger for trading volume than for stock prices. In other words, the increase of search activity is strongly associated with the increase of trading volume, but the probability that this trading raises stock prices is not so high, compared to the results provided by prior studies using the U.S. data. This difference probably derives from the fact that our sample period includes major negative economic shocks such as the 2008 world financial crisis and the 2011 Great East Japan Earthquake; also, the presence of individual investors, whose online search activity is expected to be well-associated with stock trading, is smaller in Japan than in the U.S.[3] Our results indicate that the applicability of the "price pressure hypothesis" may depend on circumstances and market conditions.

[1] Da et al. (2011) provide detailed explanation on Google Trends.

[2] The data obtained from Google Trends can also be used to forecast various kinds of economic and social activities, such as automobile sales and the number of tourists (Choi and Varian 2009) and the number of patients suffered from influenza (Ginsberg et al. 2009).

[3] The percentage of individual shareholders in the U.S. and Japan is explained in the next section.

The rest of this article is organized as follows. Section 2 provides a literature review and background information. Sections 3 and 4 explain our methodology and data. Section 5 discusses empirical results. Concluding remarks are provided in Section 6.

2. Literature review and background information

2.1 Literature review

In contrast to traditional asset pricing models based on the efficient market hypothesis, two theories focus on the relationship between investor's attention and stock prices. The so-called "investor recognition hypothesis" states that an increase in visibility of a firm conveys new information to investors who do not possess its stocks and persuades some of them to buy them (Merton, 1987). More recent studies on behavioral finance document that public attention alone is enough to move stock prices even without any new information. The so-called "price pressure hypothesis" or "attention theory" states that individual investors tend to buy stock that attracts their attention, because individual investors do not have enough time or resources to examine thousands of stocks (Barber and Odean 2008).[4] Attention is an important factor in selecting which stocks to buy among the huge number that are available. By contrast, individual investors do not encounter the same search problem when selling stocks they already own and therefore know.[5] As a consequence, stocks capturing investors' attention and searched intensively tend to generate abnormally high returns and trading volume.

Although it is empirically difficult to separate the effects of the two hypotheses, several empirical studies have shown the relationship between investors' attention and stock price movements. Early empirical studies examine the relationship between stock price movements and their appearance in the mass media.[6] For instance, Huberman and Regev (2001) report that a mention in the *New York Times* of a small biotechnology company could generate a positive effect on its stock prices. Fehle et al. (2005) find significantly positive stock price reactions for firms that appeared in the ads of Super Bowl broadcasts during 1969-2001. Takeda and Yamazaki (2006) document that stock prices of companies that appeared in the Japanese TV program "Project X" tended to increase after the broadcast. Fang and Peress (2009) investigate the relationship between

[4] It is important to note that this hypothesis applies only to individual investors but not necessarily to institutional investors. More detailed explanation is given in the next section.

[5] It is possible that individual investors conduct intensive search when they sell short. However, Baber and Odean (2008; p.786) state that individual investors seldom sell short.

[6] Other measures are extreme returns. For instance, Barber and Odean (2008) show that individual investors are net-buyers of stocks in news and high trading volume.

media coverage and the stock returns, showing that stocks without media coverage generated higher returns than those with high media coverage. Likewise, Kim and Meschke (2011) provide evidence that CEO interviews on CNBC tended to increase stock prices between 1999 and 2001.

Other studies have focused more on the relationship between online search activity and stock-trading behavior. Antweriler and Frank (2004) find that the online messages posted on Yahoo! Finance and Raging Bull are associated with stock returns and trading volume. Rubin and Rubin (2010) show that the frequency at which firm-related articles are edited in Wikipedia is negatively correlated with forecast errors by analysts and their forecast dispersions.

More recent studies have used the aggregate search frequency in Google, namely, the Search Volume Index (SVI), as a measure of investor attention (Da et al. 2011; Joseph et al. 2011; Bank et al. 2011; Vlastakis and Markellos 2012). The weekly SVI, which is provided by Google for the search term public via Google Trends, is the number of searches for the term scaled by its time-series average. Da et al. (2011) and Joseph et al. (2011) use ticker symbols of stocks as search keywords in Google Trends and show that online ticker searches are positively associated with abnormal returns of stocks included in Russell 3000 for the period 2004-2008 and those included in S&P500 for the period 2005-2008, respectively.

Unlike Da et al. (2011) and Joseph et al. (2011), Bank et al. (2011) and Vlastakis and Markellos (2012) employ the company name instead of the ticker symbol and get the same results as the studies using the ticker symbol. Bank et al. (2011) report that an increase in search queries tends to raise stock liquidity and trading activity in the German market. Similarly, Vlastakis and Markellos (2012) find that demand for information is positively related to volatility and to trading volumes for 30 major stocks traded in the NYSE and NASDAQ.

Da et al. (2011) and Joseph et al. (2011) discuss two potentially important problems arising from the use of the company name. First, users may perform searches for a company name to obtain a variety of information that may not be directly associated with investing decisions (e.g. product information, location and hours of stores, recruitment, sports teams sponsored by the company, scandals, and so on). Second, there are potentially many different ways to spell the name of a company. To overcome the first problem, Vlastakis and Markellos (2012) assume that components irrelevant to investment purposes are either random or systematically caused by seasonality or time trends whose effects can be removed by appropriate pre-processing of the data. They also use keywords with the largest search volume based on Google Insights.

2.2 Japanese setting

To the best of our knowledge, no paper has examined the relationship between search intensity and stock-trading behavior in the Japanese market. We set out to investigate whether the "price pressure hypothesis" would hold true for the Japanese market. For this purpose, we gather data between 2008 and 2011, a period that saw some significant outside influences on Japanese stock prices. These considerations and others relating to the Japanese market are detailed below.

First, the presence of individual investors is lower in Japan than in the U.S. According to the Tokyo Stock Exchange (2013), individual shareholders have accounted for approximately 20% of total shareholders for more than a decade. This percentage is less than the 37-40% ratio of individual investors in the U.S. stock market during 2005-2012 reported by the Board of Governors of the Federal Reserve System (2013). Barber and Odean (2008) state that their hypothesis applies only to individual investors, not to institutional investors. This is partly because institutional investors may face a search problem when selling stocks among the many they hold. In addition, attention may not be so rare a resource to institutional investors, who devote more time to searching and use more advanced technology than do individual investors. Thus, the smaller percentage of individual investors in Japan may weaken the linkage between the search intensity and the increase of stock prices.

Second, the linkage between the search intensity and the increase of stock prices may also be weakened by our choice of the more recent sample period than that of prior studies. Our sample period includes the 2008 world financial crisis and the 2011 Great East Japan Earthquake. These major negative economic events may have induced individual investors to search for news of companies whose stocks they hold to examine whether those stocks are worth keeping or selling off. In other words, the search intensity may have been associated not only with buying stocks but with selling them.

Third, the Japanese security code is expressed as a four-digit number and is not suitable for a search keyword. Thus, we use the company name as a search keyword. Considering the fact that different investors may search the same firm using several variations of its name, we also use abbreviated names based on the selection process specified in the next section.

3. Data

3.1. Keyword selection

In the present study, we use the online search intensity to measure the degree of investors' interests. The data on the online search intensity is obtained from Google Trends (http://www.google.co.jp/trends/). Google Trends provides a time-series data, called SVI, on the frequency of searches of specific keywords at the time and location specified by the user. The weekly SVI for a keyword is the number of searches for that keyword scaled by its time-series average, such that it takes a number between 0 and 100.

It is worth to note that the SVI is based on the number of searches of a specific keyword among searches of all keywords conducted in the same period. In other words, the SVI shows the relative frequency of searches. The relative here has two meanings: cross-sectional and time-series. First, the SVI does not increase when the number of searches of a specific keyword is less than that of other keywords. Second, the SVI for a specific week may vary across a given period of time, because the SVI takes the value of 100 when the number of searches is the highest in the period specified by the users. In addition, if the weekly search frequency is extremely low, we cannot obtain it or acquire a monthly data instead. Furthermore, Da et al. (2011, p.1467) discuss the possibility that SVIs on the same keyword may be slightly different when they are downloaded at different points of time,[7] because Google calculates SVI from a random subset of the historical search data. We downloaded the data in September, 2012, and used it unaltered.

In the present study, we examine the relationship between investors' interests and stock price movements in Japan. Our sample consists of 189 stocks included in the representative index, Nikkei 225, which is a simple average of 225 stocks that are actively traded in the first section of the Tokyo Stock Exchange. One thing to note is that the Japanese market does not employ the ticker symbols used in the US markets. Thus, we use the keywords derived from company names.

As discussed in the previous section, Da et al. (2011) and Joseph et al. (2011) identify two potential problems arising from the use of a company name. First, a user who searches for a company name may be seeking information irrelevant to investment purposes. Second, there are many different ways to spell the name of a company. To address these two issues, we first follow the assumption made by Vlastakis and Markellos (2012), stating that components irrelevant to investment purposes are random. To deal with the second issue, we employ multiple keywords for a company based on the following selection procedure.

---

[7] According to Da et al. (2011), the correlations of SVIs downloaded at different points of time are more than 97%.

1. We exclude "Kabushiki kaisha (Co., Ltd., Inc., etc.)" and "Holdings" from the keywords.
2. We add several abbreviations of the company name based on the following three steps.

2-1. We make a list of abbreviations of the company name stated in Wikipedia (http://ja.wikipedia.org/wiki/).[8]

2-2. We check SVIs for the abbreviations and the company name and delete them from the list if their SVIs are below 1/10 of the highest SVI.[9]

2-3. We exclude the abbreviation that can be used for too general a meaning.[10]

3. We add "- (minus) name of sports or sports team" to keywords when we find a name of a sporting topic or sports team in the related keywords of Google Trends.[11]
4. We add "- (minus) something irrelevant to the company" to keywords when we find such irrelevant things in the related keywords of Google Trends.

With regard to other search conditions, we use the data on "all categories" of "web search" conducted in "Japan" (location) for the period between "January, 2008 and December, 2011." We download the weekly data for 208 sample periods during these four years.

3.2. Sample selection

Our samples consist of the firms included in Nikkei 225 as of September 14, 2012. Table 1 shows our sample selection process. Specifically, we delete firms in the following four categories: (1) Firms whose stock prices cannot be obtained for January, 2008 and December, 2011 from Toyo Keizai's Kabuka CD-ROM 2012 because of mergers and acquisitions or being established during the sample period. (2) Firms whose SVIs are 0 for more than five weeks. (3) Firms for which all keywords are excluded based on the keyword selection procedure described above. (4) Firms whose SVIs are less than 10 for more than 105 weeks, about half of the 208 sample weeks. For these firms SVIs reach 100 after a dramatic increase for some period due to a big

8 It is worth to note that Wikipedia can be edited by anybody and its contents are not always reliable. However, we need the abbreviations commonly used and accepted. For this purpose, we believe that abbreviations stated in Wikipedia are suitable for our research.

9 Google Trends enables us to compare the average SVIs on five keywords at maximum for the specified period.

10 For example, Ubekosan (Ube Industries, Ltd.) has three abbreviations: Ube, Kosan, and Ubeko. Kosan (Industries) is too general and thus excluded from the keywords.

11 For example, "Toray – Arrows" gives the results on the search frequency for information related to Toray Industries, Inc. but not to the volleyball team "Toray Arrows."

accident or event, and are relatively low for the other periods. (5) Firms in railroad, electricity, and gas industries. SVIs of these firms are 100 in the Great East Japan Earthquake in March, 2011.

(Table 1 here)

4. Methodology

4.1. Search intensity, abnormal returns, and trading volume

In the present study, we examine the relationship between the search intensity and stock-trading behavior. We first divide our sample of 189 firms into four quartiles based on three indicators of search intensity related to the SVI in the preceding week. For all three indicators, Q4 consists of the firms with the highest search intensity, while Q1 comprises those with the lowest search intensity. We then estimate abnormal stock returns based on the Fama-French three-factor model and trading volumes for each portfolio. The three indicators of search intensity are presented in subsection 4.1.1, our method of estimating abnormal returns based on the Fama-French three-factor model in 4.2.2, and our method of calculating trading volumes in 4.1.3.

4.1.1 Three indicators of the search intensity

We define the average return of the portfolio Qk (k=1, 2, 3, 4) at time t by denoting n_{Qk} as the number of stocks included in the portfolio Qk:

$$R_{Qk,t} = \frac{\sum A_{i,t}}{n_{Qk}}$$

(1)

where $A_{i,t}$ is the return of the stock included in the portfolio Qk, which is based on the three models explained below.

We first define Model 1, which is based on the level of the SVI in the preceding week:

$$A_{i,t} = \left\{ R_{i,t} \,\middle|\, \frac{(k-1)N}{4} < RANK(SVI_{i,t-1}) \leq \frac{kN}{4}, \quad 1 \leq i \leq N \right\} \tag{2}$$

where $R_{i,t}$ is the return of the stock i at time t and $SVI_{i,t-1}$ is the SVI of the stock i at time $t-1$. $RANK(SVI_{i,t-1})$ is the order of the $SVI_{i,t-1}$ among sample firms. Model 1 is the same model used in Joseph et al. (2011). However, this model entails a problem when we compare multiple SVIs of different stocks. Consider the case in which the search intensity of stock *j* becomes extremely large at *t0* (SVI_{j,t_0}) because of some major news item. This makes $SVI_{j,s}$ ($s \neq t_0$) relatively low for other periods. If this happens, most of the $SVI_{j,s}$ can be classified in Q1, when comparing SVIs of other stocks not affected by such a big shock.

To avoid the problem of Model 1, we define Model 2 based on the change in SVI:

$$A_{i,t} = \left\{ R_{i,t} \,\middle|\, \frac{(k-1)N}{4} < RANK(\Delta SVI_{i,t-1}) \leq \frac{kN}{4}, \quad 1 \leq i \leq N \right\} \tag{3}$$

where $\Delta SVI_{i,t-1}$ is the change in the SVI:

$$\Delta SVI_{i,t-1} = SVI_{i,t-1} - SVI_{i,t-2} \tag{4}$$

Model 2 enables us to have many SVIs of a particular stock in the same quartile when comparing SVIs of multiple stocks. However, Model 2 also contains another problem. As explained before, Google Trends calculates SVIs based on a random subset of the historical data on search activities. This sometimes makes SVIs for the same keyword slightly different when they are downloaded at different points of time. In other words, ΔSVI can either be positive or negative, depending on the timing of the download.

To avoid the problem of Model 2, we employ Model 3 based on abnormal SVI ($ASVI$) as follows:

$$A_{i,t} = \left\{ R_{i,t} \,\middle|\, \tfrac{(k-1)N}{4} < RANK(ASVI_{i,t-1}) \leq \tfrac{kN}{4}, \quad 1 \leq i \leq N \right\} \tag{5}$$

where $ASVI_{i,t-1}$ is the difference between $SVI_{i,t-1}$ and the median of the $SVI_{i,s}$ (*s*=*t-2*,…,*t-8*) for seven periods.[12]

$$ASVI_{i,t-1} = SVI_{i,t-1} - Median(SVI_{i,t-2}, \ldots SVI_{i,t-8}) \tag{6}$$

Using $ASVI$, the indicator based on Model 3 becomes more stable than the one based on Model 2.

4.1.2 Abnormal stock returns

We next estimate abnormal returns of the portfolio *Qk* based on the Fama-French three-factor model:

---

[12] The definition of ASVI follows Da et al. (2011).

$$R_{Qk,t} - R_{f,t} = \alpha + \beta_m (R_{m,t} - R_{f,t}) + \beta_s SMB_t + \beta_h HML_t + \varepsilon_{Qk,t} \tag{7}$$

where $R_{f,t}$ is the risk-free rate at time *t*, $R_{m,t}$ is the market return at time *t*, and $(R_{m,t} - R_{f,t})$ is the risk premium at time *t*. SMB_t is the difference between simple average returns of small and big stocks based on market capitalization. HML_t is the difference between simple average returns of high and low book-to-market stocks. Following Joseph et al. (2011), we regard α as the abnormal return of stocks.

The validity and robustness of the three-factor model developed by Fama and French (1993) in the Japanese capital markets have been shown by Kubota and Takehara (2010). We obtain the data regarding the three factors from Financial Data Solutions, Inc., which calculates the three factors in the Tokyo Stock Exchange based on Kubota and Takehara (2010).

4.1.3 Abnormal trading volume

We calculate trading volume to examine its relationship with the search intensity. The trading volume of the stock *i* at time *t* is defined as follows:

$$TV_{i,t} = P_{i,t} \times V_{i,t} \tag{8}$$

where $P_{i,t}$ is the price of the stock *i* at time *t* and $V_{i,t}$ is its turnover. Then the abnormal trading volume ($ATV_{i,t}$) is defined as follows:

$$ATV_{i,t} = \frac{TV_{i,t} - TV_{i,avg}}{TV_{i,avg}} \tag{9}$$

where

$$TV_{i,avg} = \frac{\sum_{t=1}^{L} TV_{i,t}}{L} \tag{10}$$

Here, *L* is the period of our examination, which takes 207, 206, and 200 weeks for Models 1, 2, and 3, respectively. We calculate weekly ATVs for portfolios Q1 to Q4.

4.2. Search intensity, cross-sectional variation of arbitrage, and trading volume

Next, we conduct multivariate regression analysis using panel data, characterized by 189 stocks and 208 weeks, to examine whether the search intensity improves the Fama-French three-factor model. Specifically, we add the search intensity (*SENT*) as an explanatory variable to Equation (7):

$$R_{i,t} - R_{f,t} = \alpha + \beta_0 SENT_{t-1} + \beta_m (R_{m,t} - R_{f,t}) + \beta_s SMB_t + \beta_h HML_t + \epsilon_{i,t} \tag{11}$$

Corresponding to the univariate analysis, we use three variables of search intensity (*SENT*). The first variable is the logarithm of SVI at time *t-1* ($\ln SVI_{t-1}$) based on Model 1.[13] The second variable is $\Delta \ln SVI_{t-1}$ ($= \ln SVI_{t-1} - \ln SVI_{t-2}$) based on Model 2. The third variable is $ASVI'_{t-1}(= \ln SVI_{t-1} - median(\ln SVI_{i,t-2,} \dots, \ln SVI_{i,t-8})$ based on Model 3.

We then estimate the abnormal trading volume $ATV_{i,t}$ to examine whether this variable is affected by the search intensity, as predicted by the price pressure hypothesis. Similar to Equation (11), we use three variables of search intensity (*SENT*):

$$ATV_{i,t}=\beta_1 SENT_{t-1}+\beta_2 \frac{\sum_{i=1}^{N} ATV_{i,t}}{N} + \epsilon_{i,t} \tag{12}$$

where *N* is the number of all sample stocks.

Before estimating Equations (11) and (12), we conduct a Wu-Hausman test to determine the specifications of the models. For all models of Equation (11), the results of the test neither reject the null hypothesis that fixed effects are not significantly different from zero, nor support the existence of random effects. Thus, we choose to estimate Equation (11) using a pooled-data model. The results of the test reject the null hypothesis at the 1% significance level for Model 1 of Equation (12), while the results cannot reject the null for Models 2 and 3. However, none of the models support the existence of random effects. Thus, we choose to estimate Equation (12) using a fixed-effect model for Model 1 and a pooled-data model for Models 2 and 3.

5. Results

5.1. Search intensity, abnormal returns, and trading volume

Table 2 presents the Pearson correlation matrix of variables used for estimation of Equation (7). Because all correlations are less than 0.5, we do not consider the possibility of multicollinearity. Table 3 shows the regression results of Equation (7). Our focus is on the abnormal return α. Although Model 1 does not provide a clear pattern, α is the largest for Q4 among the four portfolios in Models 2 and 3. In addition, raw returns are the largest for Q4 in these two models. These results are consistent with the "price pressure hypothesis" that abnormal returns are positively related to search intensity.

(Tables 2 and 3 here)

[13] Considering the fact that the SVI can be zero, we take a natural logarithm of SVI+1.

Table 4 presents the average abnormal trading volume (ATV) for the four portfolios. For all three models, ATV is the largest in Q4. In addition, for Models 1 and 3, ATV is significantly larger for Q4 than for Q1. All these results are consistent with the notion that the search intensity is positively associated with trading volume.

(Table 4 here)

5.2. Search intensity, cross-sectional variation of arbitrage, and trading volume

Table 5 shows regression results from examining the relationship of the search intensity with cross-sectional variation of arbitrage and trading volume. Panels A and B correspond to Equations (11) and (12), respectively. Table 6 presents the Pearson correlation for the variables used for the regression of Equation (12), which indicates that there is only a low correlation between variables.

(Table 5 and 6 here)

In Panel A, the coefficient on the search intensity is significantly positive at the 10% level for Model 1, but those for Models 2 and 3 are not significantly different from zero. This means that the search intensity tends to increases stock prices, though the significance level is somewhat weak. In Panel B, for all three models, the coefficients on the search intensity are significantly positive at the 1% level. In other words, the search intensity is strongly and positively associated an abnormal trading volume. These results are consistent with our prediction in the sense that investors who search for the company tend to trade stocks of the company.

Our results presented in Panel A are only weakly support the “price pressure hypothesis.” As explained in Section 2, there are possibly three reasons to explain the difference. First, our sample period (January, 2008 and December, 2011) includes several major calamitous events such as the world financial crisis starting from the Lehman shock and the Great East Japan Earthquake. At these times, investors may have searched for the information on the companies whose stocks they already held more actively than in other periods because they wanted to know whether to sell the stocks, especially of companies and industries negatively impacted by such events. In other words, the search intensity may not have led to the purchase of stocks but rather to their sale during these negative events.

Second, the percentage of individual investors is smaller in Japan than in the U.S. According to the Tokyo Stock Exchange (2013), the percentage of individual investors has been approximately 20% in all Japanese stock exchanges in the past decade, while households accounted for 37-40% of the U.S. shareholders for 2005-2012, according to the Board of Governors of the Federal Reserve Board (2013). Because the investors who use Google Trends are likely to be individual investors, the effect of online searches may be smaller in Japan than in the U.S. Third, individual investors, who search for company information are likely to buy and sell securities on the same day. Such day trading may not be fully captured in our weekly data. Our results thus indicate that the applicability of the "price pressure hypothesis" may depend on circumstances and conditions on the market.

6. Concluding remarks

In this paper, we examine the relationship between the online search intensity and stock-trading behavior to analyze how investors' interests affect stock prices. We use Google Trends to obtain data on the online search intensity, measured as the frequency of searches of company names on Google. We employ a sample of 189 stocks included in Nikkei 225 searched between 2008 and 2011. We note that our results on the positive relationship of the search intensity with stock prices are somewhat weak compared to those on the relationship with trading volume. This difference may result from the facts that 1) our sample period includes major negative events such as the world financial crisis in 2008 and the Great East Japan Earthquake in 2011, and 2) the presence of individual investors is relatively small in Japan. Future research is needed to clarify to what kinds of circumstances and conditions the "price pressure hypothesis" does apply.

References

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

Baltagi, B.H., and Levin, D. (1992). "Cigarette taxation: Raising revenues and reducing consumption." *Structural Change and Economic Dynamics* 3: 321–335.

Bank, M., Larch, M., and Peter, G. (2011). "Google search volume and its influence on liquidity and returns of German stocks." *Financial Markets and Portfolio Management*, 25(3): 239-264.

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: 785–818.

Board of Governor of the Federal Reserve System "Financial Accounts of the United States: Flow of Funds, Balance Sheets, and Integrated Macroeconomic Accounts: Historical Annual Tables 2005-2012." available at http://www.federalreserve.gov/releases/z1/current/annuals/a2005-2012.pdf.

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

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

Fama, E. F. (1976). *Foundations of Finance: Portfolio Decisions and Securities Prices*, Basic Books: AZ.

Fama, E., and French, K. (1993). "Common risk factors in the returns on stocks and bonds." *Journal of Financial Economics* 33: 3-56.

Fang, L., and Peress, J. (2009) "Media coverage and the cross-section of stock returns." *Journal of Finance* 64(5): 2023-2052.

Fehle, F., S. Tsyplakov, and V. Zdorovtsov (2005) "Can Companies Influence Investor Behavior through Advertising? Super Bowl Commercials and Stock Returns." *European Financial Management*, Vol. 11, No. 5, 625-647.

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinsky, M. S., and Brilliant, L. (2009). "Detecting influence epidemics using search engine query data." *Nature,* 457: 1012–1014.

Kim, Y. H., and Meschke, F. (2011) "CEO interviews on CNBC." Working Paper, available at SSRN: http://ssrn.com/abstract=1745085.

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

Jagannathan, R., Kubota, K., and Takehara, H. (1998), “Relationship between labor-income risk and average return: empirical evidence from the Japanese stock market.” *Journal of Business* 71: 319-47.

Joseph, K., Wintoki, M. B., and Zhang, Z. (2011). “Forecasting abnormal stock returns and trading volume using investor sentiment: Evidence from online search.” *International Journal of Forecasting* 27: 1116 – 1127.

Kahnemanm, D. (1973) *Attention and Effort*. Prentice-Hall: Englewood Cliffs, NJ

Kubota, K. and Takehara, H. (2010) “Expected return, liquidity risk, and contrarian strategy: evidence from the Tokyo Stock Exchange.” *Managerial Finance* 36(8): 655-679.

Merton, R. (1987). “A simple model of capital market equilibrium with incomplete information.” *Journal of Finance*, 42(3): 483–510.

Rubin, A., and Rubin, R. (2010) “Informed investors and the internet.” *Journal of Business Finance and Accounting* 37: 841-865.

Takeda, F., and Yamazaki, H. (2006) “Stock Price Reactions to Public TV Programs on Listed Japanese Companies.” *Economics Bulletin* 13(7): 1-7, 2006.

Tokyo Stock Exchange (2013) “2012 Shareownership Survey.” Available at http://www.tse.or.jp/english/market/data/shareownership/index.html.

Vlastakis, N., and Markellos, R. N. (2012). “Information demand and stock market volatility.” *Journal of Banking and Finance* 36(6): 1808-1821.

Table 1: Sample selection

No. of firms included in Nikkei 225	225
(less) No. of firms without stock prices	3
(less) No. of firms whose SVIs take 0 for more than five weeks	12
(less) No. of firms whose keywords are not appropriate	4
(less) No. of firms whose SVIs take less than 10 for more than 105 weeks	5
(less) No. of firms in railroad, electricity and gas industries	12
No. of firms in our sample	189

Table 2: Pearson correlation matrix

	Model 1			Model 2			Model 3		
	LN(SVI)	$R_m - R_f$	SMB	Δ LN(SVI)	$R_m - R_f$	SMB	ASVI'	$R_m - R_f$	SMB
$R_m - R_f$	-0.006			-0.025			-0.019		
SMB	0.002	-0.384		-0.012	-0.396		0.020	-0.400	
HML	0.025	0.063	0.046	0.012	0.063	0.049	0.019	0.081	0.031

Table 3: Portfolio analysis based on Fama-French three-factor model

Portfolio	Raw Return (%)	α		$R_m - R_f$		SMB		HML	Adjusted R^2 (%)	Obs (week)
				Model 1						207
Q4	-0.15	0.0890		1.1740	***	-0.2046	***	-0.0113	95.23	
		(1.44)		(57.36)		(-4.00)		(-0.18)		
Q3	-0.05	0.1947	**	1.2526	***	-0.0469		-0.0699	93.46	
		(2.55)		(49.67)		(-0.74)		(-0.89)		
Q2	-0.16	0.0520		1.2024	***	0.0106		-0.0018	92.59	
		(0.67)		(46.76)		(0.17)		(-0.02)		
Q1	-0.10	0.1412	*	1.2105	***	-0.0776		-0.0916	92.34	
		(1.75)		(45.45)		(-1.16)		(-1.10)		
				Model 2						206
Q4	-0.02	0.2027	**	1.2007	***	-0.0756		-0.0404	92.62	
		(2.58)		(45.96)		(-1.15)		(-0.50)		
Q3	-0.13	0.0984		1.2102	***	-0.0875		-0.1038	94.64	
		(1.47)		(54.45)		(-1.56)		(-1.50)		
Q2	-0.18	0.0386		1.1807	***	-0.1136	*	-0.0139	93.99	
		(0.55)		(51.00)		(-1.94)		(-0.19)		
Q1	-0.11	0.1055		1.2650	***	0.0064		-0.0190	92.56	
		(0.08)		(0.03)		(0.07)		(0.09)		
				Model 3						200
Q4	-0.07	0.1410	*	1.1861	***	-0.0120		-0.0291	93.52	
		(1.92)		(48.84)		(-0.19)		(-0.38)		
Q3	-0.14	0.0855		1.2418	***	-0.0565		-0.0492	94.31	
		(1.18)		(52.05)		(-0.93)		(-0.65)		
Q2	-0.10	0.1247	*	1.1885	***	-0.1213	**	-0.0569	94.39	
		(1.80)		(51.99)		(-2.07)		(-0.78)		
Q1	-0.15	0.0723		1.2258	***	-0.0967		-0.0329	90.71	
		(0.77)		(39.76)		(-1.23)		(-0.34)		

Notes:

1. ***, ** and * indicate statistical significance at the 1%, 5% and 10% levels, respectively.

2. Figures in parenthesis represent t-value.

Table 4: Abnormal trading volume and search intensity

Portfolio	Abnormal trading volume		
	Model 1	Model 2	Model 3
Q4	0.032	0.017	0.030
Q3	0.002	-0.001	0.001
Q2	-0.002	-0.011	-0.014
Q1	-0.033	-0.006	-0.021
Q4－Q1	0.064 **	0.023	0.051 *
	(1.64)	(0.64)	(1.49)
Obs (week)	207	206	200

Notes:

1. ** and * indicate statistical significance at the 5% and 10% levels, respectively.

2. Figures in parenthesis represent t-value.

Table 5: Regression results

Panel A: Search intensity and cross-sectional variation in arbitrage

Model 1			Model 2			Model 3		
Intercept	-0.236		Intercept	0.114	***	Intercept	0.111	***
	(-1.22)			(5.04)			(4.85)	
LN(SVI)	0.094	*	ΔLN(SVI)	-0.015		ASVI'	0.059	
	(1.85)			(-0.15)			(0.55)	
$R_m - R_f$	1.209	***	$R_m - R_f$	1.213	***	$R_m - R_f$	1.210	***
	(161.75)			(160.96)			(160.22)	
SMB	-0.081	***	SMB	-0.068	***	SMB	-0.072	***
	(-4.32)			(-3.57)			(-3.71)	
HML	-0.044	*	HML	-0.046	*	HML	-0.039	
	(-1.88)			(-1.94)			(-1.64)	
Obs	39,123		Obs	38,934		Obs	37,800	
Effects (cross section)	None		Effects (cross section)	None		Effects (cross section)	None	
Adjusted R^2 (%)	44.63		Adjusted R^2 (%)	44.75		Adjusted R^2 (%)	45.39	

Panel B: Search intensity and abnormal trading volume

Model 1			Model 2			Model 3		
Intercept	-2.156	***	Intercept	-0.007		Intercept	-0.026	***
	(-29.26)			(-1.44)			(-5.00)	
LN(SVI)	0.565	***	ΔLN(SVI)	0.146	***	ASVI'	0.243	***
	(29.27)			(6.70)			(9.83)	
Average(ATV)	0.099	***	Average(ATV)	0.098	***	Average(ATV)	0.093	***
	(27.47)			(26.74)			(25.37)	
Obs	39123		Obs	38934		Obs	37800	
Effects (cross section)	Fixed		Effects (cross section)	None		Effects (cross section)	None	
Adjusted R^2 (%)	3.40		Adjusted R^2 (%)	1.82		Adjusted R^2 (%)	1.75	

Notes:

1. ** and * indicate statistical significance at the 5% and 10% levels, respectively.
2. Figures in parenthesis represent t-value.

Table 6: Pearson correlation matrix

		Average(ATV)
Model 1	LN(SVI)	-0.017
Model 2	Δ LN(SVI)	-0.133
Model 3	ASVI'	-0.173