Singapore Management University

Institutional Knowledge at Singapore Management University

Research Collection School Of Information Systems

School of Information Systems

Not all that glitters is gold: The effect of attention and blogs on the investors' investing behaviors

Nan HU
Singapore Management University, hunan@smu.edu.sg

Yi DONG
University of International Business and Economics, Beijing

Ling LIU
University of Wisconsin

Lee J. YAO
Loyola University, New Orleans

Follow this and additional works at: https://ink.library.smu.edu.sg/sis_research

Part of the Computer Sciences Commons, Finance and Financial Management Commons, and the Technology and Innovation Commons

Citation

Not All That Glitters Is Gold: The Effect of Attention and Blogs on Investors' Investing Behaviors

Nan Hu[1], Yi Dong[2], Ling Liu[1], and Lee J. Yao[3]

Abstract

This article investigates the relationship between a firm's visibility in blogspaces, termed *blog exposure*, and the cross-sectional stock returns. We show that blog exposure is fundamentally different from the traditional media coverage, and securities with low blog exposure earn higher returns than stocks with high blog exposure. We further illustrate that such an effect is more prominent for stocks with low institutional ownership. Contrary to traditional media coverage, the return premium associated with blog exposure cannot be explained by either the illiquidity hypothesis or the investor recognition hypothesis based on the rational-agent framework. Instead, our results suggest that blog effect can be attributed to the limited attention theory and cannot be arbitraged due to investors' self-attribution and short-sale constraints. Our research points out the importance of blogs in information dissemination, especially for the stocks with limited attention.

Keywords

blog, expected return, limited attention, self-attribution, short-sale constraints, word of mouth communication

Introduction

In this article, we investigate whether a firm's visibility within blogspaces, termed *blog exposure*, can affect security pricing even if the blog itself does not supply any authentic news. When controlling for environmental factors such as traditional media coverage and analyst coverage, and various risk factors such as size, book-to-market (BM) ratio, beta, and momentum, our results indicate that stocks with high blog exposure earn lower returns than stocks with low blog exposure. These results are more pronounced among stocks with lower institutional ownership. We also show that a portfolio composed of stocks with low institutional ownership produces sustained future abnormal returns. Such returns peak approximately 10 months after formation of the portfolio, and reverse sharply after that,

[1]University of Wisconsin, Eau Claire, USA
[2]University of International Business and Economics, Beijing, China
[3]Loyola University, New Orleans, USA

Corresponding Author:
Yi Dong, No. 10, Huixin Dongjie, Chaoyang District, Beijing, China
Email: docljy@gmail.com

displaying a long-term return reverse pattern. In addition, the abnormal returns of a portfolio composed of stocks with high institutional ownership meander around zero over the next 12 months after formation of the portfolio. We examine three plausible explanations for this return premium, two from the rational-agent framework (liquidity and investor recognition) and one from the behavior finance framework (limited attention with short-sale constraints). Surprisingly, contrary to media coverage, for which the return premium can be explained by liquidity and investor recognition based on the rational-agent framework (Merton, 1987), the return premium of blog exposure cannot be explained by that framework. Instead, our findings show that a blog exposure premium originates from the joint forces of biased disclosure of bloggers and limited attention of consumers. Such overvaluation becomes larger over time due to consumers' biased self-attribution, while the short-sale constraints prevent such a premium from being arbitraged. Our results are most consistent with the notion in Hirshleifer, Lim, and Teoh (2009). In addition, our results show that the blog exposure effect is not subsumed by traditional media exposure, as documented by Fang and Peress (2009). However, without mainstream media (e.g., news) planting a seed of discussion, blog exposure actually has no impact on security returns.

Understanding the relation between blog exposure and stock returns is very important for the marketing community and the finance community because it demonstrates how a firm's marketing strategy within blogspaces can influence elements of its finance strategy, such as cost of capital. As a form of word of mouth (WOM)[1], blogs represent the fastest growing medium of personal publishing and the newest method of individual expression and opinion on the Internet.[2] In 2004, blogs were a fairly new phenomenon with only 5 million bloggers worldwide (Wright, 2006). However, at the time of writing this article, according to www.BlogPulse.com, there were more than 126 million blogs on the World Wide Web. We believe blogs are playing a role that is as important as that of newspapers because (a) information in a blog is not a simple reflection of what is covered by traditional news. In fact, many blogs address topics that are not covered by the mainstream media at all. Blogs might either lead or follow traditional news and (b) blogs disseminate information to a much broader audience faster and with in-depth analysis. In fact, compared with other online media, blogs are viewed as more credible. In addition, compared with traditional sources, more than three quarters of respondents view blogs as moderate to very credible.

This article adds to the growing body of studies of the valuation of online WOM literature (Antweiler & Frank, 2004; Tumarkin & Whitelaw, 2001; Wysocki, 1999). Our article is also closely related to but distinct from Fang and Peress (2009)'s media coverage study that shows that, by helping to reach a broad population of investors, mass media can alleviate information frictions and affect stock price even if it does not contain authentic news. However, Fang and Peress focus on studying the impact of traditional media coverage, measured by the number of newspaper articles about a firm on its stock returns. We concentrate on examining the influence of nontraditional media coverage—blog exposure—on security pricing, while controlling for the traditional media coverage. Furthermore, our results are distinct from Fang and Peress' conclusions. The return premium of news coverage can be explained by the rational-agent framework (Fang & Peress, 2009). However, such a return premium of blog exposure can only be explained by the joint forces of investors' behavior and short-sale constraints.

Another branch of literature related to our study is the behavior finance literature that recognizes that attention influences investors' selling and purchasing behavior, and causes asset pricing deviation from its fundamental value. The underlying reason is that investors

face a formidable search problem when buying a stock. Investors address such an issue by limiting their choice sets (Barber & Odean, 2008) to those stocks that have recently caught their attention (Odean, 1999). Because investors are overconfident (Daniel, Hirshleifer, & Subrahmanyam, 1998; Odean, 1998) and biased toward self-attribution (Daniel et al., 1998), stocks over bought heavily by individual investors will enjoy short-term positive contemporaneous returns. These results emphasize the attention driven by traditional media, such as securities mentioned in the news or securities that have gone through large volume or price changes. Our study, however, focuses on the attention driven by nontraditional media, blogs.

One practical implication of our results is that a firm's visibility within blogspaces, regardless of whether the blog discussion is positive or negative, influences investors' purchase decision. As a nontraditional media, blog discussions serve the role of disseminating information that was traditionally covered by conventional information channels such as analysts' forecasts or newspapers. Such a role is especially important for stocks with limited attention. In fact, as we documented, because blog discussions do not affect the expected stock returns when there is a lack of echo from the mainstream media, companies should plant the seed of a discussion to foster the conversation within blogspaces. Marketing managers of a firm can use blogs not only to communicate more efficiently with its customers, partners, suppliers, and other stakeholders but also to work closely with finance managers to lower the cost of capital by delivering information to a broader audience in a faster manner.

The rest of this article is organized as follows. In the next section, we discuss our data collection processes, elaborate our variable definitions, and present our empirical results. In the section ''Explaining the Blog Visibility Effect,'' we discuss three possible causes of the blog exposure effect, and in the section ''Conclusion,'' we present our concluding remarks.

Data Collection and Empirical Results

Data Collection

In this article, blog visibility/exposure is defined as the extent to which a company's products or services are discussed in blogspaces. We collect such a measurement from www.BlogPulse.com using its conversation track tool. Our data collection is composed of two steps. In Step 1, we identify the brand names of products or services associated with a company by searching a company's web site, reading its financial reports, or using researchers' domain knowledge. In Step 2, using the names identified in Step 1 as keywords, for each company, we retrieve its blog visibility over time using the conversation tracker tool provided by www.BlogPulse.com. Of the total daily blogging activities traced by www.BlogPulse.com, this measurement represents the percentage of the total blog conversation related to a particular firm, its products, or its services. It measures how (and how much) a firm's current customers, potential customers, competitors, industry peers, and so on are talking about the products or services of the firm. Hence, it represents the visibility of a firm within the overall blogspaces. Figure 1 shows one example of the blog trend for Advance Auto Parts Inc.[3]

The data on a firm's media exposure were collected from Factiva. Factiva is a database of the Dow Jones and Reuters companies. It provides timely, domestic, and international information, such as articles from the Dow Jones and Reuter's newswires and *The Wall Street Journal*. This information covers market data, firm and industry news, financial
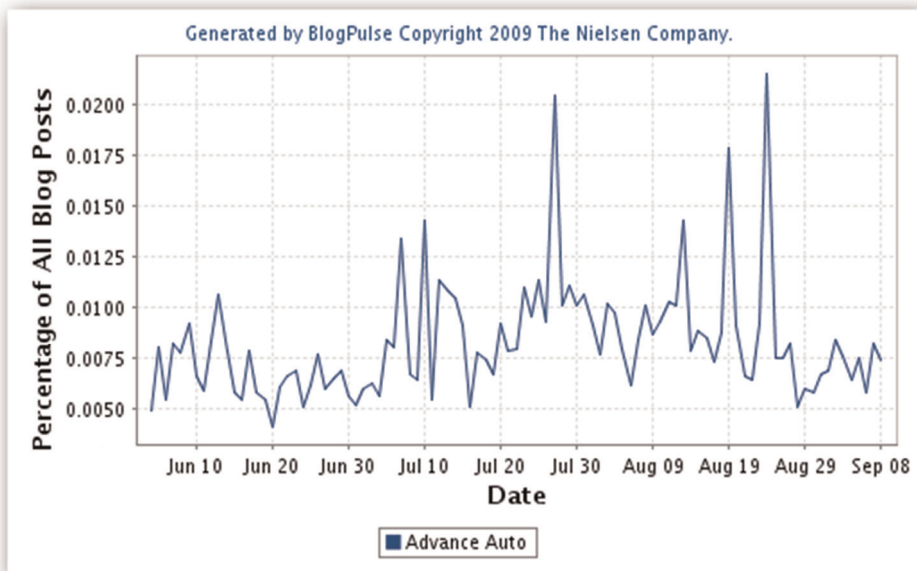
Figure 1. Blog coverage data collection
Note: In this figure, we use Advance Auto Parts Inc. as an example to show how we collect the exposure of a firm within blogspaces using www.blogpulse.com

quotes, and newspaper articles. We collected and counted all the nonredundant news items each day for the fiscal year of each company. Our analysts' forecast data were collected from the Institutional Brokers Estimates System. The company accounting data were obtained from CompuStat, and the stock return data were from the Center for Research in Security Prices (CRSP). For detailed variable definitions, please refer to the appendix.

Following Amihud (2002), we delete one firm with extreme illiquidity value. The final sample contains 404 Standard & Poor (S&P) 500 firms with daily blog visibility and traditional media coverage information from March 1, 2006, to August 22, 2006. We give detailed variable definitions in the appendix. Following Fang and Peress (2009), to minimize the noise of daily data, for each firm, we aggregate its daily blog visibility and Factiva news to a monthly level to represent its blog visibility and media coverage, respectively. Table 1 shows the summary statistics of our key variables, including blog coverage and media coverage. As we can see, only 1% of the stocks in our sample do not have media coverage, but more than 25% of the stocks in our sample do not have blog coverage. It seems that for the S&P 500 firms, and for this sample period, traditional media has broader coverage in terms of the number of stocks discussed than blog conversations. However, this does not necessarily mean that, in general, traditional news media has broader coverage than blogs.

Table 2 shows the Pearson correlation among blog coverage, media coverage, and other firm characteristics. We find blog coverage and media coverage are positively correlated, and analysts, news, and blogs have the tendency to feature the same set of stocks. This is reasonable because analysts, news, and blogs pay attention to large firms and well-known firms. Furthermore, it seems that traditional media cares more about firms with a lower

Table 1. Summary Statistics of Blog Exposure and Media (Factiva) Coverage

Blog		Factiva	
M	0.40	M	236.42
SD	1.22	SD	415.45
Skewness	5.66	Skewness	3.50
Quintile		Quintile	
100% maximum	11.89	100% maximum	2,791.00
99%	7.38	99%	2,267.00
95%	1.90	95%	1,085.00
90%	0.80	90%	535.00
75% Q3	0.22	75% Q3	224.00
50% median	0.04	50% median	95.00
25% Q1	0.00	25% Q1	41.00
10%	0.00	10%	19.00
5%	0.00	5%	7.00
1%	0.00	1%	0.00
0% minimum	0.00	0% minimum	0.00

Note: See the appendix for detailed variable definitions.

institutional ownership, whereas bloggers do not differentiate between firms with high institutional ownership and those with low institutional ownership.

Blog Coverage and Short-Term Cross-Section of Stock Return

In this section, we study whether a firm's exposure within blogspaces affects its security pricing by regressing stock returns on blog coverage, with media coverage, beta, size, BM ratio, and other risk factors controlled (Table 3, Model 1). For a robustness check, we take an incremental approach by adding momentum, media coverage, percentage of institutional ownership, and illiquidity one by one, and present the results in Models 2, 3, and 4 of Table 3, respectively. To control for the potential confounding effect caused by the difference across months and industries, we also include month fixed effect and industry fixed effect (two-digital standard industrial classification from CRSP) in all our models.

For Model 1, the coefficient of blog exposure is -0.0026 with t-value at -1.93, which is negatively associated with the following month's stock return. Therefore, stocks with high blog exposure tend to have lower stock returns compared with the stocks with low blog coverage. Such an impact is not subsumed when we add the traditional media exposure and other risk factors (Models 3 and 4). Furthermore, using Factiva as proxy for media coverage, we found that media coverage has an insignificant coefficient regardless of whether we exclude the blog exposure. The untabulated table shows that our results are qualitatively the same if we exclude those firm-month observations in which there are earning announcements for the firm in that month. Therefore, we conclude that the blog effect is not driven by month effect, industry effect, or earnings announcements.

Our media coverage results are different from those of Fang and Peress (2009). However, in their article, they focus on four influential newspapers with large subscriptions. Our media coverage includes all the newspapers included in the Factiva database. Another reason for the different results might lie in the sample selection. In their article,

Table 2. Pearson Correlation

	Blog	Factiva	Size	BM	Momentum	Dispersion	IO	Coverage	Illiquidity	Idiorisk
Blog	1									
	2353									
Factiva	0.1550	1								
	<0.0001									
	2,353	2,353								
Size	0.1334	0.5198	1							
	<0.0001	<0.0001								
	2,346	2,346	2,346							
BM	−0.1214	−0.0383	−0.2583	1						
	<0.0001	0.0685	<0.0001							
	2,261	2261	2,259	2,261						
Momentum	−0.0740	−0.0242	0.0886	−0.1616	1					
	0.0003	0.24	<0.0001	<0.0001						
	2,352	2,352	2,345	2,261	2,352					
Dispersion	0.0125	0.0050	0.0076	−0.0384	0.0218	1				
	0.5513	0.8131	0.7169	0.0722	0.3008					
	2,266	2,266	2,263	2,191	2,266	2,266				
IO	0.0088	−0.1053	0.0614	−0.0905	0.2115	−0.0500	1			
	0.6695	<0.0001	0.0029	<0.0001	<0.0001	0.0173				
	2,353	2,353	2,346	2,261	2,352	2,266	2,353			
Coverage	0.1293	0.3358	0.6064	−0.1358	0.0195	0.0012	0.0353	1		
	<0.0001	<0.0001	<0.0001	<0.0001	0.3528	0.9549	0.0916			
	2,281	2,281	2,278	2,201	2,281	2,266	2,281	2,281		
Illiquidity	−0.0212	−0.0369	−0.2677	0.0348	−0.0602	−0.0399	−0.1454	−0.1501	1	
	0.3043	0.0732	<0.0001	0.0983	0.0035	0.0578	<0.0001	<0.0001		
	2,353	2,353	2,346	2,261	2,352	2,266	2,353	2,281	2,353	
Idiorisk	−0.0038	−0.0725	−0.3569	−0.0612	0.1148	0.0035	0.0578	−0.0635	0.2501	1
	0.8558	0.0004	<0.0001	0.0036	<0.0001	0.8681	0.005	0.0024	<0.0001	

Note: BM = book to market; IO = institutional ownership. For detailed variable definitions, refer to the appendix.

they argue that the most significance of high media coverage, low return effect comes from the small and illiquid firms. However, in our article, we use only S&P 500 firms, which are all big firms.

## Blog Coverage and Long-Term Cross-Section of Stock Returns

In the previous section, we documented the short-term return premium associated with low blog exposure firms. In this section, we study the long-term return impact of blog exposure. Figure 2 represents the cumulative abnormal returns of stocks with different blog coverage, starting from Month 1 after the portfolio formation. Each month, we sort our sample into three groups according to their monthly blog exposure. Then, based on capital asset pricing model, for each stock, we estimate its abnormal return in the subsequent 13 months. Low blog curve (Figure 2) represents the average abnormal returns for the lowest blog coverage group, whereas high blog curve represents those of the highest blog coverage group. Figure 2 shows that the highest blog coverage portfolio consistently has insignificant abnormal returns starting from the 1st month after the formation of the portfolio to the next 13

Table 3. Blog Coverage, Media Coverage, and Stock Returns

	Model 1	Model 2	Model 3	Model 4
Intercept	0.0116 (−0.45)	0.0295 (−1.03)	−0.0043 (−0.15)	0.0041 (−0.12)
Blog	−0.0026 (−1.93)		−0.0033 (−2.45)	−0.0033 (−2.39)
Factiva		0.0000 (−0.87)	0.0000 (−0.65)	0.0000 (−0.65)
Beta	−0.0026 (−1.1)	−0.0028 (−1.78)	0.0003 (−0.13)	0.0002 (−0.06)
Size	0.0024 (1.72)	0.0012 (0.70)	0.0021 (1.23)	0.0019 (0.87)
BM	0.0042 (1.78)	0.0040 (1.72)	0.0012 (0.51)	0.0013 (0.56)
Momentum			−0.0323 (−5.79)	−0.0311 (−5.36)
Coverage				−0.0001 (−0.25)
IO				−0.0034 (−0.68)
Illiquidity				−0.0203 (−0.95)
Adjust_R^2	.1119	.1107	.1253	.1246

Note: BM = book to market; IO = institutional ownership; SIC = standard industrial classification. In this table, the dependent variable is the following month's stock return. The independent variables include traditional risk factors and other firm characteristics that may affect expected stock returns. Our samples include 404 distinct firms. We run pooled regression, controlling month fixed effect and industry fixed effect (2-digital SIC).
$Nextmonthre = \alpha + \beta_{beta} \times beta + \beta_{size} \times size + \beta_{BM} \times BM + \beta_{momentum} \times momentum + \beta_{IO} \times IO + \beta_{coverage} \times coverage + \beta_{illiquility} \times illiquility \times \beta_{blog} \times blog + \beta_{Factiva} \times Factiva$.
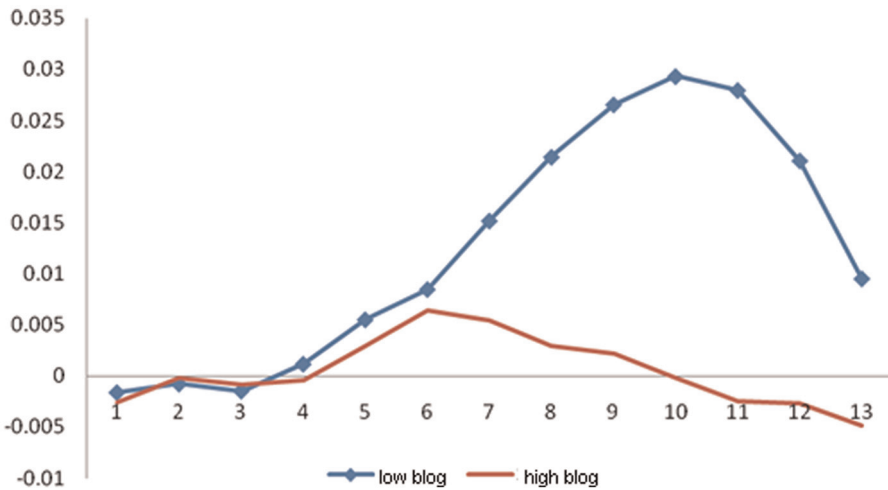
Figure 2. Cumulative abnormal return
Note: CAPM = capital asset pricing model; CAR = coverage abnormal return. Figure 2 presents the cumulative abnormal return of a high blog coverage sample and a low blog coverage sample starting from Month 1 after the portfolio formation. Each month we sort our sample into two groups according to its monthly blog coverage. Then, based on CAPM model, for each stock, we estimate its cumulative abnormal return in the subsequent 13 months. The blue line represents the average CAR for the low blog coverage group, whereas the red line plots that of the high blog coverage group.

months, and the low blog coverage portfolio enjoys a positive coverage abnormal return in the following 13 months. In addition, there is a return reversal when the cumulative abnormal returns peak at the 10th month.

Our results are not driven by the return reversals among no-coverage stocks because the abnormal returns associated with low blog visibility firms do not reverse until 10 months after the portfolio formation, while a typical return reversal pattern among no-news losers is short lived (Chan, 2003). Given that our blog coverage effect represents a long-term effect, there must be a force to prevent traders from such arbitrage. So what are the driving forces?

Explaining the Blog Visibility Effect

In this section, we examine three potential explanations of the return premium associated with blog exposure: illiquidity, investors' recognition, and bloggers' limited attention and short-sale constraints. The first two are based on the rational-agent framework, whereas the third explanation originates from behavior finance literature.

Illiquidity Hypothesis

Fang and Peress (2009) document that stocks with no media coverage enjoy higher returns than stocks with high media coverage. Furthermore, such a return premium is very stable over time. They believe that the lack of liquidity explains why such an abnormal return cannot be arbitraged. Hence, we first test whether blog visibility is similar to media coverage aroused due to illiquidity. According to the rational-agent framework, if the blog effect represents an arbitrage opportunity, it can only be persistent in the situation where some kind of impediment to trade prevents arbitrage. Hence, if the blog visibility effect is also caused by illiquidity, we expect the blog effect to be most significant in the portfolios that are composed of the most illiquid stocks.

We use multivariate regressions (either separate regressions for each portfolio or pooled together) to study whether the illiquidity can explain the cross-sectional return differences we document. Each month we sort our sample into three portfolios based on various illiquidity proxies proposed by previous literature including the Amihud (2002) illiquidity ratio, bid/ask spread, trading price, and firm size. Stocks with the highest illiquidity ratio or spread, or stocks with the lowest price or size, are the most illiquid ones. For each illiquidity proxy and each portfolio, the following month's stock returns are regressed on blog coverage with other controlled factors that are known to affect the cross-section of returns, such as beta, size, BM ratio, and momentum. To control for the potential heterogeneity across months as well as industries, we run our regression by controlling month and industry effects.

Table 4 reports the blog effect of stocks sorted by different illiquidity proxies. Due to space constraint, for each portfolio under different illiquidity measures, we report only the coefficient before blog coverage and the corresponding t statistic. If the blog visibility effect is caused by illiquidity, we expect the blog exposure effect to be most significant in the portfolios composed of the most illiquid stocks, such as stocks with the highest illiquidity and spread or those with the lowest price or size. However, the results shown in Table 4 fail to support the illiquidity hypothesis. For example, with respect to the Amihud illiquidity ratio, blog effect is significant in the low illiquidity portfolio (para = -0.0068 and t-value = -3.04) and the medium illiquidity portfolio (para = -0.0058 and t-value = -2.9), but not in the high illiquidity portfolio (para = 0.0022 and t-value = 0.67). This contradicts the illiquidity hypothesis that high impediments of trade should result in the most significant blog exposure effect. Similar examples of evidence are found when we use

Table 4. Illiquidity and the Blog Effect

	Low		Medium		High	
	Parameter	t-value	Parameter	t-value	Parameter	t-value
Illiquidity	−0.0068	−3.04	−0.0058	−2.39	0.0022	0.67
Spread	−0.0042	−1.88	−0.0048	−1.90	−0.0015	−0.55
Price	−0.0040	−1.46	−0.0079	−3.20	−0.0065	−1.68
Firm size	0.0004	0.14	−0.0015	−0.40	−0.0072	−3.57

Note: In this table, each month we classified our sample into three portfolios according to their respective illiquidity proxies. We use the Amihud illiquidity ratio (Amihud 2002), bid ask spread, size of the firm, and price to proxy for illiquidity. The following months' stock returns are regressed on blog coverage, beta, size, BM ratio, and momentum, controlling month fixed effect and industry fixed effect. For brevity, for each portfolio under different illiquidity measures, we report only the coefficient before *blog coverage* and corresponding *t* statistic.

alternative illiquidity measurements. Using price or firm size to proxy for liquidity, we observe that the blog exposure associated return premium is most significant in the portfolios composed of the most liquid stocks, for example, stocks with highest trading prices (para = −0.0065 and *t*-value = −1.68) or stocks with the biggest firm sizes (para = −0.0072 and *t*-value = −3.57). In addition, we also use bid/ask spread to proxy for illiquidity because highly illiquid stocks often have a large bid/ask spread. The coefficients of blog visibility for low and medium bid/ask spread portfolios are negative and significant but not for the high bid/ask spread portfolio (para = −0.0015 and *t*-value = −0.55)—the most illiquid stocks. Overall, our results show that the blog visibility effect disappears among the most illiquid stocks, hence are contrary to the illiquidity theory that suggests that the blog visibility effect should be the strongest for most illiquid securities.

To check the robustness of the above results, for each illiquidity proxy, we run a pooled regression using all three portfolios. To be more specific, for each illiquidity proxy, we first define one dummy variable to distinguish high illiquidity firms from low illiquidity firms. For example, the dummy variable *Rank* for Amihud illiquidity is set to one if the illiquidity ratio of that firm is higher than the median Amihud illiquidity ratio, and zero otherwise. Then, in addition to the original independent variables, we also include an interaction term between the blog coverage variable and the illiquidity dummy variable *Rank* to test whether the blog exposure effect is more significant for stocks with higher illiquidity. Our untabulated results show that consistent with the conclusions from running separate regressions, results using pooled regressions also demonstrate that illiquidity cannot explain the blog exposure effect that we have documented.

Investors' Recognition Hypothesis

Fang and Peress (2009) posit that, in addition to illiquidity, the investor recognition theory proposed by Merton (1987) can also explain the media coverage effect they observe. Under this theory, investors are assumed to have incomplete information and are aware of only a subset of the available stocks. For stocks with less investor recognition, investors with incomplete information will require higher returns as a compensation for the undiversified risk and market clearing. Therefore, we hypothesize that if blog coverage can increase a stock's recognition, then the blog effect should be more prominent for stocks with a low degree of investor recognition. Several measures, including analyst coverage, idiosyncratic

Table 5. Investor Recognition and the Blog Effect

	Low parameter	t-value	Medium parameter	t-value	High parameter	t-value
Idiorisk	−0.0088	−3.60	−0.0069	−3.18	0.0004	0.16
Coverage	−0.0012	−0.42	−0.0037	−1.31	−0.0077	−3.37
Expenditure	0.0035	0.23	−0.0068	−2.33	−0.0012	−0.39
Investor base	−0.0040	−1.47	−0.0061	−2.50	−0.0016	−0.53

Note: In this table, each month we sort our sample into three portfolios according to different investor recognition proxies, including analyst coverage, idiosyncratic risk, advertising expenditure, and the number of investors of a stock. The following months' stock returns are regressed on blog visibility, beta, size, BM ratio, and momentum as well as month fixed effect and industry fixed effect. For brevity, for each portfolio under different illiquidity measures, we report only the coefficient before *blog coverage* and the corresponding *t* statistic.

risk, advertising expenditure, and the number of investors of a stock, are adopted as proxies for investors' recognition. Analyst coverage is selected because investors listen to analysts to make investment decisions, and stocks with high analyst coverage are assumed to have a high degree of investor recognition. Idiosyncratic risk is adopted as another investor recognition measure because it represents the imperfect diversification driven by a lack of investors' recognition (AHXZ 2006). Furthermore, following previous literature (Grullon, Kanatas, & Weston, 2004; Singh, Faircloth, & Nejadmalayeri, 2005), we develop two additional investor recognition measures: advertising expenditure and the number of common shareholders of a security.

Following the methods in the illiquidity hypothesis test section, we examine whether the blog effect can be explained by investors' recognition using a separate regression approach and a pooled regression approach. For the separate regression methods, again we sort our sample into three portfolios in each month according to different investor recognition measures. Then, for each investor recognition proxy and each portfolio, following months' returns are regressed on blog visibility while controlling for other factors that are known to affect the cross-section of returns such as beta, size, BM ratio, and momentum (Table 5). Similarly, as in the section ''Illiquidity Hypothesis,'' we also ran pooled regressions to study the impact of investor recognition measures on stock returns and reach very similar conclusions as when we followed a separate regression approach.

Table 5 presents the results of separate regressions for portfolios with different degrees of investor recognition. Following Table 4, for each investor recognition proxy, we report only the coefficient and *t*-value of blog coverage in each portfolio. As we can see, when investor recognition is measured by idiosyncratic risk, the coefficient of blog visibility is significant only for the low and medium idiosyncratic risk portfolios, but not for the high idiosyncratic risk portfolio. In addition, the absolute value of the coefficient of the low idiosyncratic risk portfolio is bigger than that of the medium idiosyncratic risk portfolio. This indicates that the blog effect is more significant in portfolios with high investor recognition, failing to support the hypothesis that the blog exposure effect can be explained by investors' recognition. We reach a similar conclusion when we use analyst coverage, advertisement expenditure, and number of investors as proxies for investor recognition. The blog effect is much more prominent for stocks with higher analyst coverage, medium advertisement expenditure, and a medium investor base. None of these evidences support the investor recognition hypothesis.
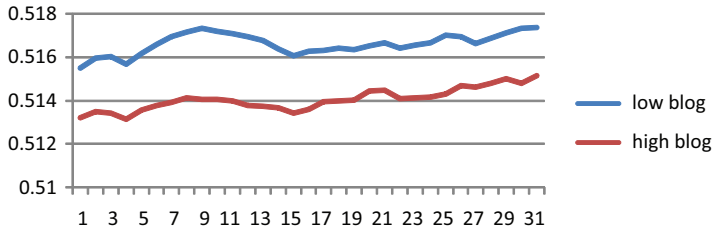
Figure 3. Percentage of buy over time

Note: This figure presents the percentage of buy transaction for firms with the low daily blog coverage during the subsequent 30 days after portfolio formation. We sort our sample into 2 portfolios based on their daily blog coverage, and low (high) daily blog coverage firms include those whose daily coverage is lower (higher) than the median blog coverage on that day. We also conduct a similar analysis by sorting our sample into 4 or 10 portfolios and reach a similar conclusion.

Limited Attention Hypothesis and Short-Sale Constraints

Based on the rational-agent framework, the illiquidity hypothesis and the investors' recognition hypothesis are able to explain how news media coverage drives the cross-section of stock returns (Fang & Peress, 2009). However, such a rational-agent framework cannot justify the blog exposure effect we have documented in the section ''Data Collection and Empirical Results.'' Given that the blog exposure effect persists over time, there must be other forces driving this return premium and preventing it from being arbitraged.

In this section, we propose that the limited attention theory (Hirshleifer et al., 2009) offers an explanation for the cross-sectional return differences we document, while the short-sale constraints sustain such return premium over time. From a disclosure prospective, blogs serve as a positively biased disclosure channel. Investors with limited attention will selectively interpret the biased disclosure by assuming that no news is good news. This will make investors net buyers of low blog exposure stocks. The biased interpretation behavior of investors on blog disclosure is consistent with the framework proposed by Hirshleifer et al. (2009) about how investors interpret the disclosures of those firms with limited attention. To make matters worse, investors with limited attention are overconfident (Daniel et al., 1998) and biased toward self-attribution (Daniel et al., 1998). In other words, individual investors believe they are better in assessing blog information (i.e., overconfident) and they selectively trust the messages in WOM and selectively validate their beliefs (i.e., biased toward self-attribution). The overall net buying behavior will be sustained over an even longer period of time.

Are investors net buyers of low blog exposure securities? To prove that consumers are more likely to be net buyers of securities with low blog exposure, we estimate the percentages of buy transactions out of the daily total number of transactions for a high blog exposure portfolio and a low blog exposure portfolio, respectively (Figure 3). We follow the buy and sell classification algorithm proposed by Lee and Ready (1991), which is commonly used in early literature, such as Easley, Kiefer, O'Hara, and Paperman (1996), and Easley, Hvidkjaer, and O'Hara (2002). Following Lee and Ready (1991), a transaction is defined as a buy (sell) if it is executed above (below) the midpoint of the bid and ask price. For trades on the bid/ask midpoints, we use a ''tick test'' to determine whether it is a buy or sell. To be more specific, a trade is a buy (sell) if it is executed at a higher (lower) price than the previous trade. For those trades that have the same price as the previous trade, we look at the historical price

Table 6. Short-Sale Constraints and the Blog Effect

Variable	Low IO parameter	t-value	Medium IO parameter	t-value	High IO parameter	t-value
Blog	−0.0067	−2.58	−0.0036	−1.40	−0.0020	−0.82

Note: IO = institutional ownership; BM = book to market. In this table, in each month, we sort our sample into three portfolios according to their short-sale constraints proxy, the institutional ownership. Then the following month's stock returns are regressed on blog visibility, beta, size, BM ratio, and momentum, as well as month fixed effect and industry fixed effect. For brevity, for each portfolio, we report only the coefficient and t-value of blog coverage

until we find a change in the trade price. Following Lee and Ready, for the computation, we match our trade prices with 5-s-old quotes. After classifying the buy and sell for each trade, we cumulate the number of buys and sells in a day to get the aggregated daily number of buy and sell. Then the daily percentage buy is defined as the total number of buy transactions divided by the total number of all transactions on that day.

Figure 3 presents the percentage of buy transactions for firms with low daily blog coverage for the 30 days subsequent to portfolio formation. We sort our sample into 2 portfolios based on their daily blog coverage. Low (high) daily blog coverage firms include those whose daily coverage is lower (higher) than the median blog coverage on that day. We also conduct a similar analysis by sorting our sample into 4 or 10 portfolios and reach a similar conclusion. Our results show that, regardless of the magnitude of blog exposure, on average, there is an increase in the percentage of buy transactions after the portfolio formation, which might be driven by the overall market situation. Furthermore, Figure 2 shows that, compared with the high blog exposure portfolio, the low blog exposure portfolio has a higher percentage of buy transactions.[4] The untabulated mean difference comparison also validates this conclusion (difference = 0.0022 and t-value = 3.99). In addition, this difference remains relatively stable over time after portfolio formation. In addition, the untabulated results show that most of the increase in the percentage of buy transactions concentrates on small investors.

Short-sale constraints. In our previous section, we documented that because investors with limited attention become net buyers of low blog exposure securities, stock prices go up. How about institutional investors or other rational investors? Why do they not come in and fix this "irrational" behavior? We believe that the short-sale constraints might be one of the answers. Short-sale constraints can prevent the arbitrage from happening, hence making the return pattern we observe last longer.[5] We expect, if short-sale constraints can explain the blog exposure effect, that effect should be concentrated in a portfolio composed of stocks with the highest short-sale constraints. Following Asquith, Pathak, and Ritter (2005), we use institutional ownership of a firm as a proxy for its short-sale constraints, and a firm with high institutional ownership is treated as the one with low short-sale constraints.

Table 6 shows that the blog effect is significant only in low institutional ownership groups (para = −0.0067 and t-value = −2.58), representing a high sale constraints situation. Such evidence supports our hypothesis that the blog effect we documented can be explained by short-sale constraints.

Relationship Between Blog Exposure and Media Coverage

Our readers might question how securities with low blog exposure get attention. If nobody ever talks about a stock, how can investors be aware of it? Hence, in this section, we study

Table 7. Relationship Between Factiva and Blog

Factiva	Low parameter	t-value	Medium parameter	t-value	High parameter	t-value
Blog	0.0001	0.03	0.00	−0.9700	−0.00426	−2.1800

Note: We sort our sample into three portfolios in each month according to their media coverage, proxy by Factiva. Then the following month's stock returns are regressed on blog visibility, beta, size, BM ratio, and momentum as well as month fixed effect and industry fixed effect. For brevity, for each portfolio, we report only the coefficient and t-value of blog coverage.

where the attention comes from. We believe that blog exposure needs to be built on news attention. Hence, we investigate whether our blog effect will be subsumed under the recently documented media effect (Fang & Peress, 2009). We first sort our sample into three portfolios according to their monthly media coverage, proxied by Factiva. Then for each portfolio, we regress the following month's stock returns on blog coverage, size, beta, BM ratio, and momentum. Table 7 shows that, with traditional media coverage controlled, the blog effect is significant only in high media coverage portfolios and is insignificant in low and medium media coverage portfolios. Our interpretation is that blog conversations are not a simple reflection of the information content of traditional media; hence, the blog effect cannot be fully subsumed by it. As a result, to really force the market to listen to blog conversations, mainstream media need to plant seeds to spark the discussion; otherwise, a simple blog conversation without a mainstream echo will have no market response. Furthermore, even with heavy traditional media coverage, blog conversations will not be buried and will still stimulate market responses.

Conclusion

In this article, we investigate the relation between blog coverage and the cross-sectional stock returns. We show that blog coverage is different from the traditional media coverage documented by previous literature (Fang & Peress, 2009). We find that high blog coverage is associated with low stock returns, even when controlling for other risk factors and traditional media coverage. We further illustrate that such an effect is more prominent for stocks with low institutional ownership and cannot be explained by either the illiquidity hypothesis or the investor recognition hypothesis, which have been shown in explaining the cross-sectional relation between media coverage and expected stock returns. Our interpretation is that the blog coverage effect is caused by the selective interpretation of investors with limited attention on the blog posting. The abnormal returns associated with the blog exposure effect are sustained over time and cannot be arbitraged within a short period of time due to short-sale constraints. All these things make blogs an important information dissemination channel.

However, we should carefully interpret our results because there may be some other mechanism that might cause the same phenomena we documented. For example, if overall blog contents are negative instead of positive, then ''no news is good news'' might actually be a rational response in the blog coverage context. If investors with limited attention fail to understand such a relation, then investors with limited attention are likely to be pessimistic about firms with low blog coverage compared with rational investors, which will lead to undervaluation in the current period. If stocks with low blog coverage are undervalued in the current period, subsequently they outperform stocks with high blog exposure. Even

though for our data set, our untabulated sentiment mining results rule out such an explanation,[6] future research might look deep into this issue with new data set covering big, medium, and small firms, and across a longer period of time to draw more insights.

Our study offers great insights as to the importance of the marketing activities of a firm on its expected return. Our results justify the buzz value creation of marketing strategy on firm's valuation. Our study also has policy implications for government agencies, such as the Securities and Exchange Commission because it brings to the forefront the effect and importance of blog information in the market valuation of firms. This is especially important for those firms with many small and naïve investors, who have limited channels to access and limited capabilities to process/digest value-relevant information.

Appendix
Variable Definitions

Variable	Definitions
Coverage	Analyst coverage is from the summary statistics database of Institutional Brokers Estimates System (I/B/E/S). It is based on the variable Numest (number of estimates), which represents the total number of estimators covering the company for the fiscal period (annual forecast only).
Beta	Used to control market risk, run capital asset pricing model regression using previous 60-month data.
Blog	Monthly blog coverage, defined as a sum of daily blog coverage.
Book to market	Log of (book value/market value).
Dispersion	Analyst forecast dispersion, from the summary statistics database of I/B/E/S database. It is based on stdev/medest, where stdev represents the standard deviation of the forecast and medest represents the median estimation of forecast for the fiscal period (annual forecast only).
Expenditure	Advertising expense from CompuStat Merged Fundamental Annual File. It is based on the variable XAD.
Factiva	Monthly media coverage, defined as a sum of daily newspaper news, from Factiva.
Idiorisk	Idiosyncratic risk, follow AHXZ (2006).
Illiquidity	Follow Amihud (2002) method.
Investor base	Number of common shareholders (CSHR)/ordinary shareholders of a stock. It is based on CSHR collected from CompuStat.
Institutional ownership (IO)	IO from Thomson Reuters. IO is defined as all the shares held by institutional investors divided by shares outstanding. It is on quarterly basis. Each month, we use the IO of previous quarter as the IO for that month. We also match monthly IO with the nearest quarterly end IO. Qualitatively, both measures result in similar results.
Momentum	Previous 12-month cumulative return.
Size	Log of market value, defined as PRC $\times$ SHROUT based on Center for Research in Security Prices' (CRSP) monthly file, where PRC is the closing price and SHROUT is the common shares outstanding.
Spread	Monthly bid ask spread (ask high minus bid low).
Volume	Monthly trading volume based on CRSP monthly file; it is the sum of the trading volumes during that month.

Notes

1. Blog content is written by authors (also known as bloggers). It is a form of word of mouth. Typically, there are websites comprising blog posts that are organized into categories and sorted in reverse chronological order. Most blogs allow readers to comment on individual blog posts.—Adopted from Wright (2006).
2. www.blogpulse.com. BlogPulse is an automated trend discovery system for blogs. BlogPulse applies machine-learning and natural-language processing techniques to discover trends in the highly dynamic world of blogs.
3. We admit that there might be some noise in the blog exposure proxy, depending on the types and the number of blogs www.BlogPulse.com covered. However, given the large number of blogs www.BlogPulse.com covered, we believe the blog exposure as measured by www.BlogPulse.com is representative of the overall discussion on blogspaces. Furthermore, each security's blog exposure as retrieved from www.BlogPulse.com might either underestimate or overestimate the true volume of blog conversation, depending on the keywords we specify as well the algorithm www.BlogPulse.com uses to identify the information related to a firm. However, we believe the results of such an estimate noise are more likely to be biased against our findings.
4. We should be aware that there is another mechanism that might result in the return difference between the high blog exposure stocks and the low blog exposure stocks. Investors are net sellers of the securities with high blog exposure. However, Figure 2 results rule out such a possibility. If that were true, the portfolio composed of high blog exposure stocks should have big and prolonged negative returns, but that is not the case. The portfolio return of high exposure stocks is, in fact, meandering around zero.
5. Following Fang and Peress (2009), in our article, we use investor recognition to refer to the model of Merton (1987) under the more traditional rational-agent framework, whereas the limited attention refers to the behavior financial model (e.g., Barber & Odean, 2008). Even though one of the key assumptions of the investor recognition hypothesis (tested in the section ''Investors' Recognition Hypothesis'') is that investors know about only the subset of the available stocks, Merton's (1987) model is built on the rational-agent framework and is different from the limited attention hypothesis we study here. In Merton's model, attention grabbing by itself will not influence an investor's purchase decision, whereas in our case, it does.
6. Our unreported sentiment mining results of the blog contents collected from LexisNexis database show that overall blog contents are dominated by positive sentiments.

References

Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets, 5*, 31-56.

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

Asquith, P., Pathak, P., & Ritter, J. (2005). Short interest, institutional ownership, and stock returns. *Journal of Financial Economics, 78*, 243-276.

Barber, B., & 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.

Chan, W. (2003). Stock price reaction to news and no-news: Drift and reversal after headlines. *Journal of Financial Economics, 70*, 223-260.

Daniel, K. D., Hirshleifer, D., & Subrahmanyam, A. (1998). Investor psychology and security market under- and overreactions. *Journal of Finance, 53*, 1839-1885.

Easley, D., Hvidkjaer, S., & O'Hara, M. (2002). Is information risk a determinant of asset returns? *Journal of Finance, 47*, 2185-2221.

Easley, D., Kiefer, N. M., O'Hara, M., & Paperman, J. B. (1996). Liquidity, information, and infrequently traded stocks. *Journal of Finance, 51*, 1405-1436.

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

Grullon, G., Kanatas, G., & Weston, J. P. (2004). Advertising, breadth of ownership, and liquidity. *Review of Financial Studies, 17*, 439-461.

Hirshleifer, D., Lim, S. S., & Teoh, S. H. (2009). Driven to distraction: Extraneous events and underreaction to earnings news. *Journal of Finance, 63*, 2287-2323.

Lee, C. M. C., & Ready, M. J. (1991). Inferring trade direction from intraday data. *Journal of Finance, 46*, 733-746.

Merton, R. C. (1987). A simple model of capital market equilibrium with incomplete information. *Journal of Finance, 42*, 483-510.

Odean, T. (1998). Are investors reluctant to realize their losses? *Journal of Finance, 53*, 1775-1798.

Odean, T. (1999). Do investors trade too much? *American Economic Review, 89*, 1279-1298.

Singh, M., Faircloth, S., & Nejadmalayeri, A. (2005). Capital market impact of product marketing strategy: Evidence from the relationship between advertising expenses and cost of capital. *Journal of the Academy of Marketing Science, 33*, 432-444.

Tumarkin, R., & Whitelaw, R. (2001). News or noise? Internet postings and stock prices. *Financial Analyst Journal, 57*, 41-51.

Wright, J. (2006). *Blog marketing*. New York, NY: McGraw-Hill.

Wysocki, P. D. (1999). *Cheap talk on the web: The determinants of postings on stock message boards* (Working Paper). Ann Arbor: University of Michigan.