Guru Dreams and Competition:
An Anatomy of the Economics of Blogs

Yi Dong, Massimo Massa, and Hong Zhang†

Abstract

The rise of social media has encouraged guru dreams because of their low entry barrier and highly skewed distribution of rewards (public attention). The pursuit of guru status, however, may be achieved through information provision or cheap talk, and competition inherent to social media may incentivize participants to either process better information or resort to more extreme options. Using a unique dataset of blogs covering S&P 1500 stocks over the 2006-2011 period, we find evidence that blog tones can predict future stock returns. However, competition distorts opinions rather than encourage better processing of information, leading to more exaggerated negative tones. Tests based on plausibly exogenous variations in competition introduced by the exiting of bloggers confirm this finding. Our results suggest that competition may distort the incentives for informed social media participants to supply information.

Keywords: Blogs, Social media, Information provision, Competition.

JEL Codes: G30, M41

†Yi Dong is with School of Accountancy, Shanghai University of Finance and Economics. Email: dong.yi@mail.shufe.edu.cn. Massimo Massa is at INSEAD. Email: massimo.massa@insead.edu. Hong Zhang is with PBC School of Finance, Tsinghua University, and INSEAD. Email: zhangh@pbcsf.tsinghua.edu.cn. We thank seminar participants at University of International Business and Economics, Shanghai University of Finance and Economics, and the 2015 Asian Bureau of Finance and Economics Research for their helpful comments.

Introduction

One of the most interesting phenomena of the last decade has been the rise in the popularity of social media. Unlike traditional media, social media are characterized by a low barrier to entry and very high potential for rapid public diffusion. Indeed, the Internet allows almost anyone who can use web-based technology to express his or her opinions. Any individual can, for instance, create a blog at nearly zero cost and use it to express opinions on almost anything, ranging from stock valuation and political issues to fashion, culture, and so on. More importantly, the vast body of Internet users provides bloggers with a large group of potential followers. Blogging therefore allows individuals to become salient and to attract public attention in a way that is unachievable with traditional media. Nevertheless, the possibility of monopolizing public attention is concentrated in a very small fraction of bloggers—i.e., the distribution of public attention for blogs is highly skewed. These features lay out incentives for bloggers that can be loosely defined as the "dream to become a guru" (Rosen, 1981).

Two interesting questions arise. First, are "gurus-to-be" bloggers more informed than the public media? Second, given the low entry costs of blogging, how does competition affect bloggers' behavior and shape their dreams of becoming gurus? While the literature has begun to analyze the first question by focusing on different types of social media, such as Internet message boards (e.g., Tumarkin and Whitelaw 2001; Antweiler and Frank 2004, and Das and Chen 2007, and Chen et al. 2014) and Twitter (Blankespoor, Miller, and White 2014), the second question remains largely unexplored in social media. However, a potential answer to the second question is crucial to understanding the economics of media in general and of social media in particular. We aim to fill this gap by hand-collecting a unique database of blogs covering all S&P 1500 stocks over the 2006-2011 period from LexisNexis, which allows us to both extend the literature by analyzing the informativeness of blogs and contribute to it by examining how competition affects blog behavior.

We consider the two questions separately. With respect to the first question, we entertain two alternative hypotheses: if the objective of the bloggers is to attract vast public attention, bloggers may want to release superior (non-publicly available) information in order to build a long-term reputation (the *informed guru hypothesis*). Bloggers may be informed either because they are better able to process information, because they are privy to private information, or simply because they engage in collecting costly private information to reap the benefits of a better reputation. Alternatively, bloggers may be uninformed players who simply rephrase what is already published in the public media to cater to and attract public attention (the *cheap-talk hypothesis*). These different incentives can be empirically disentangled by exploring the ability of the blogs to predict the return of the stocks that are covered in the blog.

We follow Loughran and Mcdonald (2011) and use linguistic analysis to define various tone variables to analyze a blog, including *positive tone* (the sentiment score of positive words in the blog), *negative tone* (the sentiment score of negative words in the blog), *net tone* (the difference between positive and negative tone—a typical blog has both positive and negative words), and *tone extremism* (the maximum value for the positive and negative tone of the same blog article). When we aggregate these tone variables at the stock level and link them to stock return, we find that blog *net tone* helps to predict abnormal stock performance over the following month. Specifically, a one-standard-deviation increase in blog net tone is related to a 3.3% higher annualized out-of-sample DGTW abnormal stock return (the adjustment follows Daniel et al. 1997). Furthermore, positive tone and negative tone predict positive and negative DGTW returns, respectively, whereas extremism does not predict future returns.

Importantly, blog tone exhibits return predictability even after we explicitly control for the corresponding tone or net tone of the top four largest newspapers in the U.S. as well as analyst recommendations, suggesting that bloggers do disseminate information above and beyond what the public media provides. Additional tests show that hot blog coverage is positively related to informed trading and negatively related to uninformed liquidity trading. Our results therefore provide evidence in favor of the informed guru hypothesis rather than the cheap-talk hypothesis.

Once we establish the informativeness of the blogs, we move on to the second question, which centers on how competition affects bloggers' behavior and shapes their dreams to become gurus. The key feature that makes the blog industry particularly interesting is the low, almost nonexistent, cost of entry. Standard competition theory posits that competition increases the accuracy and reduces the potential biases of information (e.g., Gentzkow and Shapiro, 2006). This would imply that informed bloggers engage in a competition based on the provision of the best possible information. These considerations help us lay out the *informed competition hypothesis*.

However, theory also posits that "information producers" may have incentives to structure their reports to cater to what "information consumers" are willing to hear (e.g., Mullainathan and Shleifer, 2005). In this case, competition may incentivize bloggers to supply more *influential* as opposed to more *accurate* information. In other words, the incentive to appear "relevant" to information users will distort the type of information that is supplied. To the extent that sensational negative information influences human beings more than positive information (e.g., Skowronski and Carlston,1989; Vaish, Grossmann and Woodward, 2008) – a phenomenon referred in the psychology literature as a "negative bias"[1] – bloggers may resort to sensational information and exaggerate their negative opinions in order to attract public attention. This represents the alternative *competing-for-sensation hypothesis*.[2]

To test these two competing hypotheses, we link competition to the tone of the blog. More specifically, we proxy for competition by using a dummy variable that takes a value of one if the number of bloggers covering the firm—i.e., the competitors that a particular blogger faces—is among the top quartile in the cross section and zero otherwise. We find that the competition dummy shifts the blog *net tone* from its negative mean further in the negative direction by an additional 15%. This negative impact is both statistically significant and economically sizable. Moreover, competition significantly enhances the magnitude of *negative tone*, whereas its impact on *positive tone* appears insignificant. These results suggest that competition increases the extremism of the tone because of its impact on negative tone. They are robust to the use of alternative proxies for competition.

Overall, these findings show that bloggers are informed and that there is a positive correlation between competition among bloggers and the distortion in the tone of information, with more competition being linked to more distortion. To explore the direction of causality, we consider difference-in-differences (DiD) tests based on exogenous variations in competition introduced by the exit (i.e., shutdown) of existing bloggers. We first identify the treatment group as firms that experienced reduction in competition among bloggers covering them due to the exit/shutdown of one or more bloggers. Then, for each treated firm, we use propensity scores to find a control firm with similar characteristics either in the same industry or in the whole economy in the pre-shutdown period, except that the control firm did not experience a similar blogger exit in the post-shutdown period. We then examine whether the blogs covering the treatment group of firms in the post-shutdown period exhibit different degrees of extremism in their tones from the blogs covering the control group. Our DiD test explicitly controls for blogger, time, and firm-fixed effects, allowing us to identify the *within-blogger* change for treated firms upon the exit of other bloggers.

We find that when the degree of competition is reduced due to the exit of bloggers, the remaining bloggers respond by expressing less negative opinions. In particular, blogs covering the treatment group of firms in the post-shutdown period exhibit a less *negative tone*, whereas their *positive tone* is largely unaffected. The blog *net tone* moves toward the positive direction due to the above effects. These observations are robust regardless of whether we select control firms from the same industry or

[1] As noted by Vaish, Grossmann, and Woodward (2008), for instance, "Across an array of psychological situations and tasks, adults display a negativity bias, or the propensity to attend to, learn from, and use negative information far more than positive information." Public media may also resort to sensationalism, for instance to publish rumors that are unlikely to materialize in the future (e.g., Ahern and Sosyura 2015).

[2] This intuition is not dissimilar from the traditional wisdom that a convex payoff function encourages risk taking as a response to competition, except that bloggers take additional risk by using a more extreme tone to express the same opinion.

from the whole economy. Since a blogger typically covers multiple firms, the blogger's exit is unlikely to negatively influence the fundamentals for any particular covered firm. In this regard, reverse causality is unlikely to explain the above observations. Furthermore, industry- or characteristics-related spurious correlations (between firm operation and blogger exit) are unlikely to explain our results due to the construction of our control group as well as the list of fixed effects we use. In this regard, our findings lend support to a causal influence as discussed in the *competing-for-sensation hypothesis*.

While the above results strongly suggest that competition enhances extremism in blogs, extremism could also arise because competition incentivizes bloggers to process more extreme information, as opposed to just exaggerating more sensational negative opinions without new information. To rule out this possibility, we decompose blog tones into the part induced by competition and the part unrelated to competition (i.e., the rest). We find that the part of the blog tone that is driven by competition does not have any predictive power in terms of future stock returns. In contrast, the part of the blog tone that is unrelated to competition still exhibits significant predictive power for future returns, in terms of both net tone and negative tone. This suggests that bloggers, far from processing more extreme information, do in fact just exaggerate more sensational negative opinions without new information. In other words, competition induces extreme views that have little information content.

We finally conduct a battery of additional tests and robustness checks to provide more insights regarding our main findings. As a particular robustness check to the blogger-shutdown test, we use a different testing ground of similar spirit based on the change in the number of blog platforms. During our sample period, three new popular blog platforms started their operations in the peak years of 2007-2008—i.e., Tumblr on Feb 2007, Movable Type on Dec 2007, and Posterous on May 2008—after which the number of blog platforms stabilized. Since bloggers use blog platforms to publish blogs, the period of 2007-2008 witnessed an abnormal increase in competition from the supply side. We find that in this period, the impact of competition was significantly amplified for both negative tone and net tone, which renders the tone much more extreme. This additional test also supports the *competing-for-sensation hypothesis*.

Our results shed new light on the literature exploring how competition affects the dissemination of information in the financial market. Our findings are especially interesting in comparison with those in the literature on analysts. Both bloggers and analysts publish their opinions on firms and disseminate useful information in the market. Competition, however, seems to play a very different role in the two cases. Analyst opinions, for instance, are known to exhibit a positive bias owing to conflicts of interest (Brown, Foster, and Noreen, 1985, Stickel, 1992, Abarbanell, 1991, Dreman and Berry, 1995, and Chopra, 1998), and competition provides a solution to reduce bias and to enhance price efficiency

(Hong and Kacperczyk 2010; Kelly and Ljungqvist 2012). In contrast, conflicts of interest constitute a minimal issue for bloggers. Rather, bloggers seem to resort to negative bias to attract public attention, especially in the presence of competition.

Hence, while bloggers are incentivized to supply information in the pursuit of guru status, competition appears to distort information. This suggests that while bloggers – and social media in general – have a positive role in terms of information provision, competition weakens such a role by reducing the informational contribution of social media. The economics of social media – and in particular the part related to information provision – therefore seems to differ completely from what we have learned from the existing financial market. The general and directional negative blog bias that is induced by competition also differs from the effect of rumor or political polarization often observed in public media (e.g., Ahern and Sosyura, 2015; Groseclose and Milo, 2005).

Our work also contributes to the emerging literature on social media. While a vast body of literature has examined the impact of public media on the stock market (e.g., Barber and Loeffler 1993; Huberman and Regev 2001; Busse and Green 2001; Tetlock 2007; Engelberg 2008; Tetlock, Saar-Tsechansky, and Macskassy 2008; Fang and Peress 2009; Engelberg and Parsons 2011; Dougal et al. 2012; Gurun and Butler 2012; Solomon 2012), the impact of innovations in the domain of social media remains underexplored. The few existing studies on Internet message boards (e.g., Tumarkin and Whitelaw 2001; Antweiler and Frank 2004, and Das and Chen 2007, and Chen et al. 2014) and Twitter (Blankespoor, Miller, and White 2014) document a role of social media in disseminating information in the market. Until the present time, however, blogs—a hugely important social phenomenon—have been ignored in finance. We contribute to this literature by indicating how blogs are informed and how they can predict stock performance, which is, to the best of our knowledge, the first evidence for this specific form of social media. This evidence also extends the literature on the predictability of stock returns. More importantly, blogs allow us to explore the impact of competition on social media. Our results thus shed new light on how competition affects different sectors of the economy depending on the incentive structure of the participants.

The remainder of this paper is organized as follows. In Section II, we describe the data and the main variables that we use. In Section III, we ask whether blogs are informed. In Section IV, we link blog tone to the degree of competition among bloggers. In Section V, we assess the informativeness of blog tone due to competition. A brief conclusion follows.

II. Data and Main Variables

We collected blog information for all S&P 1500 stocks for the period from 2006 to 2011. More specifically, the LexisNexis database provides information about the identity of the bloggers, the complete text of each blog published by the blogger, the date and time of the blog posting, and the keywords of the blog. From these data, we retrieve all the blogs for which the keywords contain any of the S&P 1500 stocks. Appendix 2 provides an example of a blog.

We then apply linguistic analysis to each blog in the sample and link the outcome of the analysis to the other variables of the firm that we can identify from the CRSP/COMPUSTAT database. In addition to these databases, we obtain analyst information from I/B/E/S and newspaper articles published in The Wall Street Journal, The New York Times, Washington Post, and USA Today from LexisNexis.

Table 1 provides a snapshot of the blog coverage in our final sample. In Panel A, the first three columns report the number of S&P 1500 firms that have blog coverage and newspaper coverage, as well as the number of bloggers in each year. We see that unlike the coverage of newspapers, the coverage of blogs increases very rapidly over our sample period from 2006 to 2011, which is consistent with the gradually increasing popularity of social public networks over this period. The final two columns report the number of newspaper articles and the number of blogs in a given year. Consistent with the trend, while the number of newspaper articles remains largely constant, the number of blog articles grows explosively from a mere 3,304 in 2006 to 233,040 in 2011. These numbers indicate the importance of social public media in general and blogs in particular in the contemporaneous market.

What supports the vast growth of blog articles is the expansion of service providers supplying blog platforms through which bloggers can post their blogs. Panel B reports the launching year for some of the largest blog platforms. The importance of these platforms is reported in the next few columns – in terms of either rank or market share.[3] We can see that before 2006, two very large platforms – "Blogger" and "Wordpress" – had already been operational; however, from Panel A, we know that the entire size of the blog industry is small. The greatest change occurred in 2007 and 2008, when the two players "Tumblr" and "Posterous" were launched.

As these two players quickly captured a combined 21% of the market share, some exogenous structural changes ensued. More specifically, in these two years, the booming of blog platforms gave potential bloggers more flexibility in finding a place to express their opinions and thus attracted vast

[3] More specifically, we draw the 2009 rank from the Mashable website, the 2010 rank from the Lifehacker website, and the 2011 rank from the Webhostingsearch website. We use the different website polls in different years because no single source provides polls in each year.

numbers of new participants, dramatically raising the degree of competition among bloggers. Our later tests will use this structural shift to examine the effects of shifts in competition.

Our analysis focuses on the following variables. The first set of variables is related to the tone of the blogs. We process the linguistic content of each blog by following Loughran and Mcdonald (2011). This allows us to compute the positive and negative tone of a blog article as the weighted value of negative/positive words in the article. We denote them as $Blog\_tone\_pos_{i,k,t}$ and $Blog\_tone\_neg_{i,k,t}$, for each blog article k covering stock i in month t. Larger values for these two variables indicate a more positive and a more negative tone, respectively. If a blogger posts more than one blog article for the same firm during the same month, we take the average value of these tone variables. To rule out irrelevant articles that mention only the name of the firm, we use the relevance score provided by LexisNexis and include only the articles whose relevance score is higher than 90%.

Importantly, an article can contain both positive words and negative words and thereby can have nonzero scores for both positive and negative tone. To capture the net effect, we also compute the difference between positive and negative tone for each article, denoted as $Blog\_tone\_net_{i,k,t}$. Finally, to capture the degree of "extremism" – i.e., whether the article includes very positive or very negative words – we define the degree of extremism of the blog tone, $Blog\_tone\_extreme_{i,k,t}$, as the maximum value of the magnitude of the positive and negative tone – i.e., $\max\,(Blog\_tone\_pos_{i,k,t}, Blog\_tone\_neg_{i,k,t})$.

For the stock-level analysis, we aggregate the blogs at the stock level by averaging the values for all the relevant blogs that cover the same stock on a monthly basis. This procedure leads to a set of blog variables, $Blog\_tone\_pos_{i,t}$, $Blog\_tone\_neg_{i,t}$, $Blog\_tone\_diff_{i,t}$, and $Blog\_tone\_extreme_{i,t}$, that capture the average values for positive tone, negative tone, net tone, and degree of extremism for all the blogs covering the same stock in a given month, respectively. We define blog coverage ("$Blog\_coverage_{i,t}$") directly at the firm level as the number of blog articles that are posted on a firm in a given month.

To explore the impact of competition, we also aggregate blogs at the blogger-stock level by averaging the values for all the blogs written by the same blogger covering the same stock on a monthly basis. This procedure leads to the following variables: $Blog\_tone\_pos_{i,j,t}$, $Blog\_tone\_neg_{i,j,t}$, $Blog\_tone\_net_{i,j,t}$, and $Blog\_tone\_extreme_{i,j,t}$, which respectively define the average values for positive tone, negative tone, net tone, and degree of extremism for all the blogs written by blogger j covering stock i in month t.

We also construct and control for the corresponding newspaper tone variables by aggregating articles of the leading four newspapers at the stock level. For firm i in month t, the average positive tone, average negative tone, their difference, and the degree of extremism are labeled $News\_tone\_pos_{i,t}$, $News\_tone\_neg_{i,t}$, $News\_tone\_net_{i,t}$, and $News\_tone\_extreme_{i,t}$, respectively. Consistent with the case for blogs, only news articles with relevant scores that are above 90% are included. Newspaper coverage is also captured directly at the firm level as the number of newspaper articles that are published about a firm in a given month.

We also consider a set of firm-specific dependent or control variables. The variable *C2* is defined as in Llorente et al. (2002). It is a proxy of liquidity that measures the impact of trading volume on return autocorrelation. This variable specifies whether there is informed trading (positive *C2*) or liquidity trading (negative *C2*). The variable *Flow* measures the unexpected stock-level mutual fund flow as defined in Frazzini and Lamont (2008). *DGTW_ret* is the abnormal return following Daniel et al. (1997). It adjusts stock returns netting out the benchmark returns of portfolios made of stocks that are matched with the stocks held in the evaluated portfolio on the basis of size, book-to-market ratio, and prior-period return characteristics of the stocks.[4]

Among the control variables, *BM* is the book-to-market ratio. *Size* is the log value of a firm's total asset. *Ret* is the monthly return. *Momentum* is the previous 12-month cumulative return. *Turnover* is monthly volume turnover. *Analyst_num* refers to analyst coverage, calculated as the total number of analysts covering the firm. *Analyst_rec* refers to analyst recommendations, with the value increasing in more positive recommendation. That is, the proxy reverses the original numerical value of analyst recommendation reported in I/B/E/S using 6 minus the median recommendation in the month. Finally, *Dispersion* is the standard deviation of the analyst earnings forecast (i.e., EPS) standardized by the median analyst earnings forecast. All the variable definitions are reported in appendix A.

We lay out the descriptive statistics for the characteristics of blog and newspaper coverage in Table 2. In Panel A, we report the summary statistics for the stock-level blog and newspaper tone variables, including their entire sample mean, median, standard deviation, and quartile values at the 25th and 75th percentiles of the distribution. Panel B reports the summary statistics for the same list of blog and newspaper variables in the subsample when blog or newspaper coverage is not zero. From these two panels, we see that the bloggers typically write more articles about firms than the top four newspapers, which illustrates the importance of blogs as an economic source of information dissemination. Furthermore, when blog and newspaper coverage is nonzero, blogs are generally more positive than newspapers – i.e., blog articles have a more positive tone and a less negative tone, suggesting that the information that is delivered by blogs is also likely to differ from that provided by newspapers.

[4] See http://www.rhsmith.umd.edu/faculty/rwermers/ftpsite/DGTW/coverpage.htm for a detailed description.

Panel C reports the distribution of other firm variables, including *C2*, *Flow*, *DGTW_ret*, *BM*, *Size*, *Ret*, *Momentum*, *Turnover*, *Analyst_num*, *Analyst_rec*, and *Dispersion*. The correlation matrix among the major variables is reported in Panel D. We can see that the blog net tone is positively correlated with DGTW return and that the magnitude of negative blog tone is significantly negatively correlated with DGTW return. These observations suggest that blogs may contain useful information about stock returns.

III. Are Bloggers Informed?

Our first question is whether bloggers are informed or whether they simply rely on cheap talk to attract attention. We answer this question in two steps. First, we ask whether the market perceives bloggers to be informed, and we then directly test whether they have information.

We start by asking whether the market perceives bloggers to be informed. We expect that if blogs are informative, their presence will proxy for the presence of more informed traders and therefore fewer liquidity traders. We therefore relate the presence of blog coverage to stock characteristics that proxy for informed trading and liquidity trading. We estimate the following specification:

$$Y_{i,t+1} = \beta_0 + \beta_1 \times Blog\_coverage_{i,t} + C \times M_{i,t} + \varepsilon_{i,t+1}, \quad (1)$$

where $Y_{i,t+1}$ is, alternatively, *C2* and *Flow*, for stock i in period $t+1$; $Blog\_coverage_{i,t}$ refers to lagged blog coverage; and $M_{i,t}$ stacks a list of control variables, including newspaper coverage, *BM*, *Size*, *Ret*, *Momentum*, *Turnover*, *Analyst_num*, *Analyst_rec*, and *Dispersion*. The other variables are defined as above. We estimate a panel specification with firm- and time-fixed effect, and we cluster the standard errors at the firm level. Unreported results indicate that our results are generally robust to the use of Fama-Macbeth specifications.

The results are reported in Table 3. The first three columns report the results for *C2* and *Flow*. Recall that a positive *C2* implies informed trading, while negative *C2* implies liquidity trading (Llorente et al., 2002). We see that blog coverage increases the value of *C2*, which suggests that blog coverage is more related to informed trading than to liquidity trading. Models (4) to (6) further verify this result by replacing *C2* with uninformed mutual fund flow at the stock level. We find that blog coverage is associated with less uninformed flow. This is consistent with the notion that uninformed investors become less involved with the presence of more informed trading in the market. Overall, this table provides preliminary evidence that blogs are generally associated with information that goes above and beyond what public media—major newspapers—provide.

Next, we directly test for the informativeness of blogs by focusing on “the tone” of their content and estimating the following specification:

$$DGTW\_ret_{i,t+1} = \beta_0 + \beta_1 \times Blog\_tone_{i,t} + C \times M_{i,t} + \varepsilon_{i,t+1} \quad (2),$$

where $DGTW\_ret_{i,t+1}$ is the out-of-sample abnormal performance of stock i in month $t+1$; $Blog\_tone_{i,t}$ refers to the list of variables describing blog tone, including the signed difference between the positive tone and the negative tone of blogs (*Blog_tone_net*), the positive tone of blogs (*Blog_tone_pos*), the negative tone of blogs (*Blog_tone_neg*), and the degree to which the tone is extreme (*Blog_tone_extreme*); and $M_{i,t}$ stacks a list of control variables, including newspaper tone, *BM*, *Size*, *Ret*, *Momentum*, *Turnover*, *Analyst_num*, *Analyst_rec*, and *Dispersion*. We again include firm- and time-fixed effects, and cluster the standard errors at the stock level. Note that to conduct this test, we already aggregate blog tones at the stock level in a given month.

We report the results in Table 4. We control for analyst recommendations in each model. To highlight the extent to which blogs can provide information above and beyond public media, we also tabulate the impact of blog tone in specifications in which we control for newspaper tone. The results indicate that the difference between the positive tone and the negative tone of the blogs is highly informative. These findings hold whether we consider the base specification (Model 2) or whether we control for the degree to which the blog tone is extreme (Model 8). Furthermore, the effect is not only statistically significant but also economically relevant: a one-standard-deviation increase in *Blog_tone_net* is related to a 3.3% higher annualized DGTW return.[5]

When we decompose the tone difference into positive and negative tone, we see that positive tone predicts higher subsequent stock return and negative tone predicts lower subsequent stock return. Hence, both the positive tone and the negative tone of blog articles are generally more informative than public media. In contrast, extremism does not seem to have any predictive power for stock returns. It is important to note that the predictive power of the blogs survives even after we control for analyst recommendations and newspaper tone and that newspaper tone typically affects neither the economic magnitude nor the statistical significance of the return predictability of blogs, suggesting that blogs consist of information that is very different from what public media provides. Overall, these results support the *informed guru hypothesis*, indicating that blogs generally tend to be informed rather than focus on cheap talk.

IV. Competition and Blog Tone

Next, we examine the impact of competition on blogs. We first relate blog tone to the degree of competition in the blog market. More specifically, we estimate the following panel specification:

[5] In Model 1, we first compute the impact on monthly returns as $0.10 \times 2.69 = 0.27\%$, where 0.10 is the regression coefficient and 2.69 is the standard deviation of tone difference. We then annualize the compounded impact of 0.27% as 3.3%.

$$Blog\_tone_{i,j,t+1} = \beta_0 + \beta_1 \times Competition_{i,j,t} + C \times M_{i,j,t} + \varepsilon_{i,t+1} \quad (3),$$

where $Blog\_tone_{i,j,t+1}$ is the average tone of blogs written by blogger j covering stock i in month $t+1$, alternatively defined as either the signed difference between the positive tone and the negative tone of the blogs (*Blog_tone_net*), or the positive tone of the blogs (*Blog_tone_pos*), or the negative tone of the blogs (*Blog_tone_neg*) or the degree to which the blog tone is extreme (*Blog_tone_extreme*). The vector $M_{i,j,t}$ stacks control variables for stock i and fixed effects for blogger j. We also include time-fixed effects and cluster the standard errors at the stock level.

We report the results in Table 5. In Panel A, we use a dummy variable (*Competition_dummy*) to capture the impact of competition. The variable takes a value of one if the number of bloggers covering the firm—i.e., the competitor that a particular blogger faces—is among the top quartile and zero otherwise. In Panel B, we use a continuous variable (*Competition_con*), which is computed as the logarithm of the number of bloggers covering the firm, to proxy for competition. In both panels, we employ as the dependent variable *Blog_tone_net* in columns (1)-(3); *Blog_tone_pos* in columns (4)-(6); *Blog_tone_neg* in columns (7)-(9); and *Blog_tone_extreme* in columns (10)-(12).

We see that competition has a very significant impact on the way that blog articles are written. In Panel A, Models (1) to (3) indicate that the competition dummy typically moves the blog *net tone* further in the negative direction, with the economic magnitude of the impact being approximately 15% of its mean value.[6] Consistent with this negative impact, Models (7) to (9) document that competition increases the prevalence of negative tone in blogs. The last three models also indicate that competition increases the extremism of the tone of blogs accordingly. In contrast, competition does not seem to affect positive tone. Panel B further confirms that the impact of competition is robust when we use the continuous proxy for competition. These results provide preliminary evidence in favor of the *competing-for-sensation hypothesis*, indicating that blog tone becomes more negatively biased and extreme when competition is higher.

Interestingly, the tone of the analysts is negatively related to the blog net tone. If we decompose blog tone into positive and negative tone, we see that analysts' tone is negatively related to both positive and negative blog tone.[7] This result suggests that the tone of the blogs is very different from the tone of professional market watchers, such as the analysts.

---

[6] The economic magnitude is computed as the regression parameter of the competition dummy variable in Model 3, which is -0.11, scaled by the mean value of tone difference of -0.71.

[7] Note that a positive regression coefficient between the magnitude of negative blog tone and analyst recommendation means a negative correlation—i.e., better analyst recommendations are typically associated with more negative blog tone.

To explore the potential causal impact of competition, we consider difference-in-differences tests based on exogenous variations in competition introduced by the exit of existing bloggers. We first identify the treatment group as firms that experienced reduction in competition due to the shutdown of one or more bloggers. Then, we define the pre- and post-shutdown period as the three-month periods before and after the shutdown, respectively. For each treated firm, we use propensity scores to find a control firm with similar characteristics in the pre-event period either in a same industry or in the whole economy.

We then examine whether the blogs covering the treatment group of the firms in the post-shutdown period exhibits a different degree of extremism in tone. We adopt the following panel specification with blogger, firm, and time (year-month)-fixed effects and standard errors clustered at the firm level:

$$Blog_{tone_{i,j,t+1}} = \beta_1 \times Treat_j + \beta_2 \times Post_{j,t} + \beta_3 \times Treat_j \times Post_{j,t} + C \times M_{i,j,t} + \varepsilon_{i,t+1} \quad (4),$$

where $Blog\_tone_{i,j,t+1}$ is the average tone of blogs written by blogger j covering stock i in month $t+1$, alternatively defined as either the signed difference between the positive tone and the negative tone of the blogs (*Blog_tone_net*), or the positive tone of the blogs (*Blog_tone_pos*), or the negative tone of the blogs (*Blog_tone_neg*). $Treat_j$ is an indicator that equals one for our defined treatment firms and zero for matched control firms. $Post_{j,t}$ is an indicator for post-event observations and is zero otherwise. The vector $M_{i,j,t}$ stacks control variables for stock i and fixed effects for blogger j. The parameter of interest is β_3, the coefficient before the interaction term $Treat_j \times Post_{j,t}$. If shifts in competition originated from the shutdown of bloggers affect blog tones of the corresponding firms, we should expect this coefficient to be significant.

We employ a propensity score matching (PSM) procedure as follows: in the first stage, we model the probability of each firm experiencing a covered blogger's shutdown during this month as a function of following firm characteristics: *Analyst_rec*, *BM*, *Size*, *Ret*, *Momentum*, *Turnover*, *Analyst_num*, and *Dispersion*. Then, we match each treatment firm with a set of control firms for which the difference between the treated and the control firms' predicted probabilities is lower than or equal to 0.01 (the caliber width). In other words, the set of control firms have similar characterizes, both in terms of firm fundamentals, such as size, book-to-market, and recent performance, and in terms of analyst coverage and analyst opinions. When we further control for blogger and time fixed effects, the coefficient of the interaction term $Treat_j \times Post_{j,t}$ captures the *within-blogger* change for treated firms upon the exit of other bloggers—i.e., how bloggers respond to the exit of other bloggers in fine-toning their blogs.

The results with blogger and time fixed effects are reported in Models (1) to (4) of Table 6. The results show that the parameter of interest (β_3) is significantly positive for *net tone*, significantly negative for *negative tone*, and insignificant for *positive tone*. In other words, blogs covering the treatment group of firms in the post-shutdown period exhibit a less *negative tone*, whereas their *positive tone* are largely unaffected. The blog *net tone* moves toward the positive direction due to the above effects. This suggests that bloggers respond to a reduction in the degree of competition by expressing less negative opinions.

A remaining concern of the above observation is that, while we control for both blogger and time fixed effects, the exit of blogger itself could be somehow related to negative firm fundamentals. Even though reverse causality is unlikely in which blogger-exits *cause* firm fundamentals to deteriorate, blogger-exits could be spuriously correlated with some unobserved characteristics that can be interpreted as negative information by bloggers. For instance, firms with worse fundamentals could experience more blogger exit, leaving blog tones covering these firms more negative. Although the PSM process and the test for within-blogger variation partially control for this issue[8], we can further address the concern of spurious correlation by including firm-fixed effects to control for time-invariant firm fundamentals. Models (5) to (8) in Table 6 report the results when blogger, firm, and time-fixed effects are all controlled for. We can see that the further inclusion of firm-fixed effects does not change the sign and significance level of the parameter of interest (β_3) in all four models, suggesting that much of the concern of spurious correlation is alleviated in our tests.

V. Competition and the Quality of Blog Information

As a next step in our investigation of the information distortion hypothesis, we directly look at whether competition renders blog tone more negative because bloggers process more precise information or just because bloggers exaggerate information with a more extreme tone without providing any additional information.

We first examine the relationship between blog tone and competition in different sub-samples of stocks characterized by different degree of public information availability. We consider: the degree of analyst coverage (*Analyst_num*), governance quality (Aggarwal et al. 2009), and S&P 500 affiliation – i.e., whether the firm is included in the S&P 500 index. We report the results in Table 7. We see that competition among bloggers affects blog tone mostly in firms with high analyst coverage, better

[8] Unreported tests show that exiting bloggers are more positive *and* more negative, but do not differ from remaining bloggers in terms of net tone. In our DiD test, the potential influence of tone difference between exiting and remaining bloggers is controlled for by blogger fixed effect, whereas that for their coverage difference is controlled for by the PSM process. If the difference has any remaining influence, it should affect the results for both positive and negative tones, which we do not observe in data.

governance, and S&P 500 affiliation. In other words, competition exacerbates the negative tone of blogs, especially for stocks that are under high public scrutiny. These results support the *competing-for-sensation* hypothesis.

We then formally examine whether the link between blog tone and stock returns – i.e., the informational content of the blog – is due to the effect of competition among bloggers. To investigate this issue, we first decompose blog tone into the part due to competition ("fitted blog tone") and the part unrelated to competition ("residual blog tone"). We then relate these two orthogonal components to stock returns. More specifically, we estimate the following specification:

$$DGTW\_ret_{i,t+1} = \beta_0 + \beta_1 \times Blog\_tone\_fitted_{i,t} + \beta_2 \times Blog\_tone\_rest_{i,t} + C \times M_{i,t} + \varepsilon_{i,t+1} \quad (5),$$

which differs from Equation (2) in that we decompose $Blog\_tone_{i,t}$ into $Blog\_tone\_fitted_{i,t}$ and $Blog\_tone\_rest_{i,t}$ —the two components of blog tone that are induced by and unrelated to competition, respectively. We apply this decomposition to all four variables related to blog tone and report the results in Table 8. Specifically, we employ as the dependent variable *Blog_tone_net* in columns (1)-(3); *Blog_tone_pos* in columns (4)-(6); *Blog_tone_neg* in columns (7)-(9); and *Blog_tone_extreme* in columns (10)-(12).

We see that the component of blog tone that is driven by competition does not have any predictive power in terms of future stock returns, confirming the *competing-for-sensation* hypothesis. In contrast, the residual component of blog tone – i.e., the part of blog tone that is not linked to the discretionary effect of competition – predicts future returns in terms of both net tone and negative tone. In particular, a one-standard-deviation increase in the residual net tone (negative tone) predicts a 1.1% (1.09%) annualized abnormal return.[9] The predictive power of the residual tone variables confirms our earlier results that blog articles contain information for future returns. Overall, these results suggest that competition renders blog tone more negative not because they process more precise information but simply because bloggers exaggerate information with a more extreme tone without providing any additional information.

VI. Additional Tests and Robustness Checks

This section provides additional tests to provide further evidence and to assess the robustness of our earlier results. We first consider a different test, based on the formation of blog platforms, to revisit the

[9] Similar to Table 4, we first compute the impact on monthly returns from Model (3) as $0.10 \times 0.92\% = 0.092\%$, where 0.10 is the regression coefficient and 0.92% is the standard deviation of the residual of the fitted tone difference. We then annualize the compounded impact of 0.092% as 1.1%. Model (9) allows us to compute the impact on monthly returns as $0.07 \times 1.29\% = 0.0903\%$, where 0.07 is the regression coefficient and 1.29% is the standard deviation of the residual of the fitted tone difference. We then annualize the compounded impact of 0.0903% as 1.09%.

relationship between competition and blog tones. Then, we perform robustness checks to the difference-in-differences test based on the shutdown of existing bloggers. Finally, we assess whether the location of the blogger – i.e., whether a blogger requires user registration or is located in the U.S. – affects the inference.

We start by looking at the role played by blog platforms as changes in the number of blog platforms exogenously impact competition. Three popular blog platforms launched in 2007 and 2008. Tumblr was established in Feb 2007, Movable Type in Dec 2007, and Posterous in May 2008. The emergence of these platforms induced a vast increase in the number of bloggers in 2007 and 2008. To analyze the impact of this plausibly exogenous event, we estimate the following specification:

$$Blog\_tone_{i,t+1} = \beta_1 \times Competition_{it} + \beta_2 \times Competition_{it} \times Peak_t + C \times M_{i,j,t} + \varepsilon_{i,t+1} \quad (6),$$

where $Peak_t$ is a dummy variable that takes a value of 1 in the two years 2007 and 2008 and 0 otherwise. All the other variables are defined as before. The presence of time-fixed effects does not require us to also include the peak dummy.

We report the results in Panel A of Table 9 for *Competition_dummy* and Panel B for *Competition_con*. We see that the peak dummy amplifies the impact of competition for both negative tone and net tone. During the peak time, for instance, the impact of the competition dummy more than doubles the average net tone.[10] Competition also renders blog tone more extreme. In contrast, in line with our expectations, competition has no impact on positive tone.

Finally, we also collect information about the bloggers themselves in terms of registration and geographic location. First, bloggers can officially register on blog platforms, which in many cases allow them to have access to and comment on other blogs published in the same blog platform. Second, bloggers can be located in the U.S. and abroad. We then interact competition with these characteristics and report the results in Table 10 We find that the influence of competition remains similar between registered and nonregistered investors, suggesting that registration itself does not differentiate bloggers when they face competition. Furthermore, its influence on net tone remains the same across U.S. and non-U.S. investors, but non-U.S. bloggers actually exhibit less negative bias than U.S. bloggers. The second result is interesting, as it suggests that geographic proximity does not lean potentially more informed bloggers (i.e., U.S. bloggers on U.S. firms) toward information discovery when they face more competition. These results are largely consistent with our previous findings and our main working hypotheses.

[10] The regression coefficient of $Competition_{it} \times Peak_t$ in Panel A, for instance, is -1.02 when the dependent variable is *Blog_tone_net.* Hence, during peak years, the impact of the competition dummy on *Blog_tone_net* is -1.02, which by itself is 144% of the average value of *Blog_tone_net*.

Conclusion

In this paper, we study the economics of social media based on a unique dataset of blogs. Compared with traditional media, social media are characterized by a lower entry barrier and potentially high public attention, which allows participants to pursue guru status based on the articles that they posed. This new phenomenon leads to two important questions: Does social media attract attention via information processing or via cheap talk? Does competition intensify the incentive for information discovery or distort the tone of opinions expressed in blogs?

We document that bloggers are informed and that they are generally able to predict risk-adjusted stock performance, suggesting that social media can supply information above and beyond public media. However, competition generally leads to more exaggerated negative tone in blogs with little predictive power for stock returns, implying that competition in social media distorts information.

These results suggest that the impact of competition on the accuracy of information contained in blogs drastically differs from what we observe in other parts of the economy (e.g., analysts). Our results therefore shed new light not only on the economics of social media but also on the effect of competition on information dissemination in our economy.

References

Abarbanell, J. 1991. Do analysts' earnings forecasts incorporate information in prior stock price changes? *Journal of Accounting and Economics* 14 (2), 147-165.

Ahern K., and D. Sosyura. 2015. Rumor Has It: Sensationalism in Financial Media. *Review of Financial Studies* 28(7), 2050-2093

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

Barber, B., and D. Loeffler. 1993. The 'dartboard' column: Second-hand information and price pressure. *Journal of Financial and Quantitative Analysis* 28 (2), 273–84.

Blankespoor E., G. Miller and H. White. 2014. The role of dissemination in market liquidity: Evidence from firms' use of twitter. *The Accounting Review* 89 (1), 79-112

Brown, P, G. Foster, and E. Noreen. 1985. Security analyst multi-year earnings forecasts and the capital market, American Accounting Association: Sarasota, FL.

Busse, J., and T. Green. 2001. Market efficiency in real time. *Journal of Financial Economics* 65 (3), 415–437.

Chen H., P. De, Y. Hu and B. Hwang. 2014. Wisdom of crowds: The value of stock opinions transmitted through social media, *Review of Financial Studies* 27(5), 1367-1403

Chopra,V. 1998. Why so much error in analysts' earnings forecasts? *Financial Analysts Journal* 54 (6), 30-37.

Daniel, K., M. Grinblatt, S. Titman, and R. Wermers. 1997. Measuring mutual fund performance with characteristic-based benchmarks. *Journal of Finance* 52 (3), 1035-1058.

Das, S., and M. Chen. 2007. Yahoo! for Amazon: Sentiment extraction from small talk on the web. *Management Science* 53 (9), 1375–1388.

Dreman, D, and M. Berry. 1995, Analyst forecasting errors and their implications for security analysis, *Financial Analysts Journal* 51 (3), 30-42.

Dougal, C., J. Engelberg, D. Garcia, and C. Parsons. 2012. Journalists and the stock market. *Review of Financial Studies* 25 (3), 639-79.

Engelberg, J. 2008. Costly information processing: Evidence from earnings announcements. Working paper, University of California, San Diego.

Engelberg, J., and C. Parsons. 2011. The causal impact of media in financial markets. *Journal of Finance* 66 (1), 67–97.

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

Frazzini, A., and O. Lamont. 2008. Dumb money: mutual fund flows and the cross-section of stock returns. *Journal of Financial Economics* 88 (2), 299-322.

Gentzkow, M. and J. Shapiro. 2006. Media bias and reputation, *Journal of Political Economy* 114 (2), 280-316.

Groseclose, T, and J. Milyo. 2005, A measure of media bias, *Quarterly Journal of Economics* 120 (4), 1191–1237.

Gurun, U., and A. Butler. 2012. Don't believe the hype: Local media slant, local advertising, and firm value. *Journal of Finance* 67 (2), 561–598.

Hong, H. and M. Kacperczyk, 2010. Competition and bias. *Quarterly Journal of Economics* 125 (4), 1683-1725.

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

Kelly, B. and A. Ljungqvist, 2012. Testing asymmetric-information asset pricing models. *Review of Financial Studies* 25 (5), 1366-1413.

Llorente, G., R. Michaely, G. Saar and J. Wang. 2002. Dynamic volume-return relation of individual stocks. *Review of Financial Studies* 15 (4), 1005-1047.

Loughran, T. and B. McDonald. 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. J*ournal of Finance* 66 (1), 35-65.

Mullainathan, S. and A. Shleifer, 2005, The market for news. *American Economic Review* 95 (4), 1031–1053.

Rosen, S. 1981. The Economics of Superstars. *American Economic Review* 71 (5), 845–858.

Solomon, D. 2012. Selective publicity and stock prices. *Journal of Finance* 67 (2):599–638.

Skowronski, J. and D. Carlston. 1989. Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin*, 105(1), 131−142.

Stickel, S. 1992. Reputation and performance among security analysts. *Journal of Finance*, 47 (5), 1811-1836.

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

Tetlock, P., M. Saar-Tsechansky, and S. Macskassy. 2008. More than words: Quantifying language to measure firms' fundamentals. *Journal of Finance* 63(3), 1437-1467.

Tumarkin, R., and R. Whitelaw. 2001. News or noise? Internet postings and stock prices. *Financial Analysts Journal* 57 (3), 41–51.

Vaish A., T. Grossmann, and A. Woodward. 2008. Not all emotions are created equal: the negativity bias in social-emotional development, *Psychol Bulletin* 134 (3), 383-403.

Appendix A Variable Definitions

Variable Name	Variable Definitions
Blog Related Variables	
Blog_coverage	The number of blog articles that covered the firm in a month.
Blog_tone_pos	The average value of the positive tone of all blog articles that covered the firm in a month. An article's positive tone is constructed as the weighted frequency of words representing positive tone. Both the weighting scheme and the list of words of positive tone are provided in Loughran and Mcdonald (2011). The weight of a word for an article is determined by (1) the relative frequency of the word to the number of all words in that article, and (2) the relative frequency of documents containing that word to the total number of documents.
Blog_tone_neg	The average value of the negative tone of all blog articles that covered the firm in a month. An article's negative tone is constructed as the weighted frequency of words representing negative tone. Both the weighting scheme and the list of words of negative tone are provided in Loughran and Mcdonald (2011). The weight of a word for an article is determined by (1) the relative frequency of the word to the number of all words in that article, and (2) the relative frequency of documents containing that word to the total number of documents.
Blog_tone_net	The signed difference between the positive tone and the negative tone of blogs that covered the firm in a month.
Blog_tone_extreme	The maximum value of the magnitude of positive tone and that of negative tone of blogs that covered the firm in a month.
Competition_dummy	A dummy that takes a value of one if the number of bloggers covering the firm—i.e., the competitor that a particular blogger faces—is among the top quartile.
Competition_con	The logarithm of the number of bloggers covering the firm.
Age	The age of the blogger, which is the number of months from the first time the blogger appeared in the database until the current month.
Peak	A dummy variable that takes a value of 1 in the two years 2007 and 2008 and 0 otherwise.
Newspaper-Related Variables	
News_coverage	The number of news articles that covered the firm in a month.
News_tone_pos	The average value of the positive tone of all news articles that covered the firm in a month. An article's positive tone is constructed as the weighted frequency of words representing positive tone. Both the weighting scheme and the list of words of positive tone are provided in Loughran and Mcdonald (2011). The weight of a word for an article is determined by (1) the relative frequency of the word to the number of all words in that article, and (2) the relative frequency of documents containing that word to the total number of documents.
News_tone_neg	The average value of the negative tone of all news articles that covered the firm in a month. An article's negative tone is constructed as the weighted frequency of words representing negative tone. Both the weighting scheme and the list of words of negative tone are provided in Loughran and Mcdonald (2011). The weight of a word for an article is determined by (1) the relative frequency of the word to the number of all words in that article, and (2) the relative frequency of documents containing that word to the total number of documents.
News_tone_net	The signed difference between the positive tone and the negative tone of news articles that covered the firm in a month.
News_tone_extreme	The maximum value of the magnitude of positive tone and that of negative tone of news articles that covered the firm in a month.
Other Main Variables	

C2	A variable from Llorente et al. (2002) that measures the impact of trading volume on return autocorrelation.
DGTW_ret	Abnormal returns following Daniel et al., (1997), in which we adjust stock returns by the benchmark returns constructed from the portfolios that are matched with the stocks held in the evaluated portfolio based on the size, book-to-market ratio, and prior-period return characteristics of the stocks.
Flow	Unexpected stock-level mutual fund flow based on Frazzini and Lamont (2008).
Control Variables	
Analyst_num	Analyst coverage, calculated as the total number of analysts that covered the firm.
Analyst_rec	Analyst recommendations, with a larger value referring to a better recommendation.
BM	Book-to-market ratio.
Dispersion	The standard deviation of analyst earnings forecast (i.e., EPS) standardized by the median analyst earnings forecast.
Momentum	Previous 12-month cumulative return.
Ret	Monthly return.
Size	The log value of a firm's total assets.
Turnover	Monthly volume turnover.

Appendix B Example of Blog Article

Below is an example from LexisNexis, by DCist blog about firm Archstone-Smith (NYSE:ASN).

Old Convention Center Plans Finalized

BYLINE: dcist_sommer

LENGTH: 475 words

Nov. 21, 2006 (DCist delivered by Newstex) -- UPDATE: We've now gotten word from intrepid boy reporter Kriston Capps that the D.C. Council's Committee on Education, Libraries and Recreation voted to table Bill 16-734, in a motion brought by At-Large Councilmember Carol Schwartz, which carried 3 to 2 with Marion Barry, Schwartz and surprise vote Vincent Gray against Kathy Patterson and Phil Mendelson. What does this mean for the future of Williams' library plan? Hard to say. Tabling a bill is usually a pretty good way to kill it without technically doing so, but it's certainly conceivable that incoming Mayor Adrian Fenty, who has expressed his support for the new library in general terms, could resurrect his own version of the plan at a later time. For now it seems those in favor of preserving the Mies building can rest easy for a while longer, though allow us to be the first to chime in that the pressing issue at hand -- the fact that this city desperately needs an improved main public library (not to mention all the will-they-ever-open-again branches still in limbo) -- ought to be a top priority for the new mayor and council.

Condo developer Archstone-Smith (NYSE:ASN) and real estate firm Hines announced that their development plan for the old convention center site has received approval. From the press release: The approval was granted by the District of Columbia Deputy Mayor's Office for Planning and Economic Development, on behalf of Mayor Anthony Williams, and follows an intensive community outreach process which commenced in July 2005. Through public meetings with diverse stakeholders and community design workshops, input to the proposed master plan was received from more than 20 organizations. These organizations included Advisory Neighborhood Commissions 2C and 2F, the Downtown Cluster of Congregations, the Committee of 100 on the Federal City, the D.C. Chamber of Commerce, the Greater Washington Board of Trade, the Penn Quarter Neighborhood Association, the Sierra Club and the Downtown D.C. Business Improvement District.

With construction anticipated to begin in 2008, the project will include 275,000 square feet of retail space, 300,000 square feet of office space, 772 condo and other housing units, and 1900 parking spaces. You can check out more photos and details of the plan here. What do you think?

The District has also reserved approximately 110,000 square feet of land on the site that includes the location of a new central library. As we write this, the D.C. City Council is meeting to mark up Bill 16-734, the "Library Transformation Act of 2006," Mayor Williams' plan to lease out the current Martin Luther King Jr. Memorial Library building, designed by famed modernist architect Ludwig Mies van der Rohe, and construct a new central library facility at the old convention center site.

Table 1 Time Series Blog Coverage and Blog Platform

This table presents the time series summary statistics for blogs and large blog platforms. Panel A reports, for each year, the number of S&P 1500 firms that have blog coverage and newspaper coverage, the number of distinct bloggers, the number of newspaper articles and the number of blog articles. Panel B reports launching years for some of the largest blog platforms and information on their rankings and market shares. We draw the 2009 rank from the Mashable website, the 2010 rank from the Lifehacker website, and the 2011 rank from the Webhostingsearch website. We employ different website polls in different years since no single source provides polls for each year. Our sample covers the period from 2006 to 2011.

Panel A: Yearly Breakdown of Sample

Year	# of firms with blog coverage	# of firms with newspaper coverage	# of bloggers	# of newspaper articles	# of blog articles
2006	653	634	206	7,004	3,304
2007	1,093	639	747	6,986	16,739
2008	1,270	638	1,530	6,249	34,005
2009	1,366	599	1,882	5,276	67,177
2010	1,428	576	2,066	4,616	144,735
2011	1,415	537	2,195	3,843	233,040

Panel B: Major Blogger Launching Year and Rank Information

Launch Year	Blog Platform	2009 Rank	2010 Rank	2010 Lifehacker Poll	2011 Rank
1999	Blogger	2	2	16.60%	5
2003	Wordpress	1	1	55.42%	1
2004	SquareSpace		5	3.32%	
2005	Livejournal	5			
2007	Movable Type				3
2007	Tumblr	4	3	13.11%	2
2008	Posterous	3	4	8.29%	4
	Others			3.26%	

Table 2 Summary Statistics for the Main Variables

This table presents summary statistics for our main variables. Panel A reports summary statistics for blog coverage, blog tone, newspaper coverage, and newspaper tone. Panel B reports summary statistics for blog (media) coverage and blog (media) tone in the subsample where the firm-month has been covered by at least one blog (media) article. Panel C reports summary statistics for other variables used in our regressions analyses. Panel D reports Pearson correlations among empirical variables. Appendix A provides detailed definition for all variables.

Panel A

Variable	StdDev	Mean	Median	Lower Quartile	Upper Quartile
Blog_coverage	3.53	1.15	0	0	1
News_coverage	0.48	0.09	0	0	0
Blog_tone_net	1.41	-0.18	0	0	0
News_tone_net	1.19	-0.14	0	0	0
Blog_tone_pos	0.97	0.39	0	0	0
News_tone_pos	0.44	0.06	0	0	0
Blog_tone_neg	1.73	0.57	0	0	0
News_tone_neg	1.42	0.2	0	0	0
Blog_tone_extreme	1.21	0.48	0	0	0.26
News_tone_extreme	0.87	0.13	0	0	0

Panel B

Variable	StdDev	Mean	Median	Lower Quartile	Upper Quartile
Sample with Blog coverage					
Blog coverage	5.77	4.42	2	1	5
Blog_tone_diff	2.69	-0.71	-0.31	-1.19	0.45
Blog_tone_pos	1.4	1.48	1.14	0.55	1.98
Blog_tone_neg	2.82	2.19	1.44	0.72	2.68
Blog_tone_extreme	1.77	1.83	1.38	0.78	2.31
Blog_tone_conflict	0.27	0.42	0.42	0.23	0.61
Sample with Newspaper coverage					
Media coverage	1.3	1.67	1	1	2
News_tone_diff	4.46	-2.59	-1.11	-3.43	-0.31
News_tone_pos	1.53	1.12	0.58	0.00	1.57
News_tone_neg	4.93	3.71	1.82	0.68	4.84
News_tone_extreme	2.89	2.41	1.32	0.49	3.31
News_tone_conflict	0.31	0.28	0.19	0.00	0.51

Panel C

Variable	StdDev	Mean	Median	Lower Quartile	Upper Quartile
C2	0.28	-0.01	0	-0.04	0.03
Flow	32.24	-3.63	-1.66	-16.71	10.85
DGTW_ret	9.61	0.25	-0.05	-5.04	5.13
BM	0.49	0.59	0.46	0.29	0.72
Size	1.52	14.51	14.35	13.42	15.44
Ret	0.12	0.01	0.01	-0.06	0.07
Momentum	0.45	0.12	0.07	-0.15	0.31
Turnover	18.6	24.95	19.54	12.63	30.96
Analyst_num	6.92	9.71	8	4	14
Analyst_tone	0.64	3.54	3	3	4
Dispersion	0.17	0.04	0.024	0.01	0.06

Panel D Pearson Correlation Table													
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*	*(11)*	*(12)*	*(13)*
DGTW_ret	1												
(1)													
Flow	-0.01	1											
(2)	0.002		1										
C2	0.011	0.001											
(3)	0.001	0.863											
Blog_coverage	-0.011	0.018	-0.003	1									
(4)	0.001	<.0001	0.317										
News_coverage	-0.011	0.014	0.003	0.17	1								
(5)	0.001	<.0001	0.364	<.0001									
Blog_tone_net	0.007	-0.013	0.002	-0.21	-0.156	1							
(6)	0.054	<.0001	0.625	<.0001	<.0001								
News_tone_net	0.007	-0.016	-0.002	-0.136	-0.679	0.161	1						
(7)	0.032	<.0001	0.638	<.0001	<.0001	<.0001							
Blog_tone_pos	-0.004	0.025	-0.01	0.431	0.123	-0.033	-0.084	1					
(8)	0.216	<.0001	0.001	<.0001	<.0001	<.0001	<.0001						
News_tone_pos	-0.008	0.004	0.003	0.153	0.709	-0.108	-0.578	0.103	1				
(9)	0.023	0.249	0.297	<.0001	<.0001	<.0001	<.0001	<.0001					
Blog_tone_neg	-0.009	0.025	-0.009	0.436	0.193	-0.739	-0.172	0.681	0.146	1			
(10)	0.011	<.0001	0.006	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001				
News_tone_neg	-0.008	0.013	0.002	0.153	0.743	-0.158	-0.95	0.096	0.76	0.177	1		
(11)	0.014	<.0001	0.48	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001			
Blog_tone_extreme	-0.007	0.027	-0.011	0.46	0.177	-0.497	-0.148	0.87	0.137	0.946	0.156	1	
(12)	0.036	<.0001	0.001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001		
News_tone_extreme	-0.008	0.009	0.003	0.16	0.763	-0.151	-0.88	0.101	0.859	0.176	0.978	0.157	1
(13)	0.013	0.008	0.354	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	

Table 3 Impact of Blog Coverage on Informed Trading

This table presents results for the following regression model with firm and time fixed effects, and with standard errors clustered at the firm level:

$$Y_{i,t+1} = \beta_0 + \beta_1 \times Blog\_coverage_{i,t} + C \times M_{i,t} + \varepsilon_{i,t+1},$$

where $Y_{i,t+1}$ refers to *C*2 and *Flow*, for stock i in month $t+1$. *C*2 measures the impact of trading volume on return autocorrelation, and is constructed as in Llorente et al. (2002). *Flow* measures unexpected stock level mutual fund flow, and is constructed as in Frazzini and Lamont (2008). $Blog\_coverage_{i,t}$ refers to the lagged blog coverage, and $M_{i,t}$ stacks a list of control variables, including newspaper coverage. Appendix A provides definition for all variables. The superscripts ***, **, and * refer to the 1%, 5%, and 10% levels of statistical significance, respectively. The sample includes 96,428 firm-month observations over the 2006-2011 period.

	Dependent Variable = *C*2			Dependent Variable = *Flow*		
Variables	(1)	(2)	(3)	(4)	(5)	(6)
Blog_coverage	0.08		0.08	-0.04		-0.04
	(2.50)**		(2.50)**	(-2.13)**		(-2.15)**
Newscoverage		0.00	-0.01		0.07	0.08
		(0.00)	(-0.09)		(0.67)	(0.73)
Lagged Flow				0.94	0.94	0.94
				(430.97)***	(430.44)***	(430.92)***
BM	-0.46	-0.45	-0.46	0.55	0.54	0.55
	(-1.10)	(-1.08)	(-1.10)	(3.28)***	(3.26)***	(3.27)***
Size	0.22	0.23	0.22	0.74	0.74	0.74
	(0.62)	(0.64)	(0.62)	(4.70)***	(4.67)***	(4.70)***
Ret	-0.07	-0.06	-0.07	-0.30	-0.30	-0.30
	(-0.07)	(-0.06)	(-0.07)	(-0.83)	(-0.84)	(-0.83)
Momentum	-0.23	-0.23	-0.23	0.04	0.04	0.04
	(-0.82)	(-0.82)	(-0.82)	(0.35)	(0.35)	(0.35)
Turnover	0.00	0.00	0.00	0.00	0.00	0.00
	(-0.40)	(-0.15)	(-0.39)	(-0.21)	(-0.47)	(-0.24)
Analyst_num	0.08	0.08	0.08	-0.03	-0.03	-0.03
	(2.81)***	(3.00)***	(2.81)***	(-2.05)**	(-2.22)**	(-2.04)**
Dispersion	0.13	0.14	0.13	0.05	0.04	0.05
	(0.26)	(0.27)	(0.26)	(0.19)	(0.17)	(0.19)
Constant	-3.81	-3.76	-3.82	-10.07	-10.07	-10.05
	(-0.72)	(-0.70)	(-0.72)	(-4.26)***	(-4.26)***	(-4.25)***
Firm Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	96,428	96,428	96,428	95,861	95,861	95,861
R-squared	0.03	0.03	0.03	0.93	0.93	0.93

Table 4 Impact of Tone on DGTW Adjusted Return

This table presents results for the following regression model with firm and year-month fixed effects, and with standard errors clustered at the firm level:

$$DGTW\_ret_{i,t+1} = \beta_0 + \beta_1 \times Blog\_tone_{i,t} + C \times M_{i,t} + \varepsilon_{i,t+1},$$

where $DGTW\_ret_{i,t+1}$ is the out-of-sample abnormal return of stock i in month $t+1$. We construct abnormal stock returns following Daniel et al. (1997) and adjust stock returns by the returns of the benchmark portfolio matched with stock *i* in month *t*+1 on size, book-to-market ratio, and lag return.

$Blog\_tone_{i,t}$ refers to the list of variables describing blog tone, including the signed difference between the positive tone and the negative tone of blogs (*Blog_tone_net*), the positive tone of blogs (*Blog_tone_pos*), the negative tone of blogs (*Blog_tone_neg*), and the degree to which the blog tone is extreme (*Blog_tone_extreme*), and $M_{i,t}$ stacks a list of control variables, including newspaper tone. Appendix A provides definition for all variables. The superscripts ***, **, and * refer to the 1%, 5%, and 10% levels of statistical significance, respectively. The sample includes 87,442 firm-month observations over the 2006-2011 period.

	Dependent Variable = *DGTW_ret*							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Blog_tone_net	0.10	0.10					0.11	0.11
	(2.64)***	(2.54)**					(2.67)***	(2.61)***
News_tone_net		0.06						-0.05
		(1.07)						(-0.53)
Blog_tone_pos			0.14	0.14				
			(2.56)**	(2.53)**				
News_tone_pos				-0.13				
				(-0.63)				
Blog_tone_neg			-0.11	-0.11				
			(-2.99)***	(-2.89)***				
News_tone_neg				-0.03				
				(-0.51)				
Blog_tone_extreme					-0.03	-0.03	0.02	0.02
					(-0.89)	(-0.82)	(0.46)	(0.50)
News_tone_extreme						-0.10		-0.15
						(-1.69)*		(-1.31)
Analyst_rec	0.27	0.27	0.27	0.27	0.27	0.27	0.27	0.27
	(3.29)***	(3.28)***	(3.29)***	(3.27)***	(3.30)***	(3.28)***	(3.29)***	(3.27)***
BM	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36
	(1.40)	(1.40)	(1.40)	(1.40)	(1.40)	(1.41)	(1.40)	(1.40)
Size	-4.30	-4.30	-4.30	-4.30	-4.29	-4.28	-4.30	-4.30
	(-20.73)***	(-20.73)***	(-20.77)***	(-20.78)***	(-20.68)***	(-20.68)***	(-20.76)***	(-20.77)***
Ret	0.56	0.56	0.55	0.55	0.58	0.58	0.56	0.56
	(1.22)	(1.21)	(1.21)	(1.21)	(1.27)	(1.26)	(1.22)	(1.21)
Momentum	0.19	0.19	0.19	0.19	0.20	0.20	0.19	0.19
	(1.39)	(1.38)	(1.38)	(1.37)	(1.45)	(1.43)	(1.39)	(1.38)
Turnover	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02
	(-4.45)***	(-4.41)***	(-4.46)***	(-4.42)***	(-4.51)***	(-4.46)***	(-4.46)***	(-4.43)***
Analyst_num	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(-0.22)	(-0.23)	(-0.23)	(-0.23)	(-0.22)	(-0.22)	(-0.22)	(-0.22)
Dispersion	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17
	(0.52)	(0.52)	(0.52)	(0.52)	(0.53)	(0.53)	(0.52)	(0.52)
Constant	63.23	63.21	63.26	63.23	63.08	63.06	63.24	63.22
	(20.84)***	(20.84)***	(20.86)***	(20.86)***	(20.78)***	(20.79)***	(20.85)***	(20.86)***
Firm Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	87,442	87,442	87,442	87,442	87,442	87,442	87,442	87,442
R-squared	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04

Table 5 Competition Among Bloggers

This table presents results for the following regression model with blogger and time fixed effects, and with standard errors clustered at the firm level:

$$Blog\_tone_{i,j,t+1} = \beta_0 + \beta_1 \times Competition_{i,j,t} + C \times M_{i,j,t} + \varepsilon_{i,t+1},$$

where $Blog\_tone_{i,j,t+1}$ is the average tone of blogs written by blogger j covering stock i in month $t+1$, defined alternatively as the signed difference between the positive tone and the negative tone of blogs (*Blog_tone_net*), the positive tone of blogs (*Blog_tone_pos*), the negative tone of blogs (*Blog_tone_neg*), and the degree to which the blog tone is extreme (*Blog_tone_extreme*). In addition, $M_{i,j,t}$ stacks a list of control variables for stock i, fixed effects for blogger j and year-month fixed effects. Panel A uses *Competition_dummy*, which takes a value of one if the number of bloggers covering the firm—i.e., the competition that a particular blogger faces—is among the top quartile. Panel B uses the continuous value of competition, *Competition_con*, which is computed as the natural logarithm of the number of bloggers covering the firm. $M_{i,j,t}$ stacks a list of control variables. Appendix A provides definition for all variables. The superscripts ***, **, and * refer to the 1%, 5%, and 10% levels of statistical significance, respectively. The sample includes 47,660 firm-month observations over the 2006-2011 period.

Panel A

	Blog_tone_net		*Blog_tone_pos*		*Blog_tone_neg*		*Blog_tone_extreme*	
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Competition_dummy	-0.12	-0.11	0.03	0.03	0.15	0.14	0.09	0.09
	(-2.01)**	(-2.10)**	(1.32)	(1.54)	(2.61)***	(2.85)***	(2.89)***	(3.08)***
Analyst_rec		-0.38		-0.06		0.33		0.13
		(-5.57)***		(-2.27)**		(5.03)***		(3.85)***
BM		-0.06		0.00		0.05		0.02
		(-2.53)**		(-0.46)		(2.40)**		(1.84)*
Size		1.10		0.30		-0.80		-0.25
		(5.61)***		(3.91)***		(-4.11)***		(-2.27)**
Ret		0.38		0.12		-0.27		-0.08
		(6.30)***		(4.92)***		(-4.65)***		(-2.41)**
Momentum		0.00		0.00		0.00		0.00
		(-3.84)***		(-1.58)		(3.29)***		(2.29)**
Turnover		0.01		0.00		-0.01		0.00
		(2.43)**		(0.23)		(-2.46)**		(-2.08)**
Ananlyst_num		0.09		0.01		-0.08		-0.03
		(2.42)**		(0.53)		(-2.29)**		(-1.77)*
Dispersion		-0.04		-0.03		0.01		-0.01
		(-0.41)		(-0.70)		(0.12)		(-0.18)
Constant	-2.76	-2.22	0.75	0.81	3.50	3.03	2.13	1.92
	(-4.51)***	(-3.30)***	(1.25)	(1.32)	(9.40)***	(6.06)***	(5.41)***	(4.30)***
Blogger Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	47,660	47,660	47,660	47,660	47,660	47,660	47,660	47,660
R-squared	0.46	0.47	0.38	0.38	0.51	0.51	0.50	0.50

Panel B								
	Blog_tone_net		*Blog_tone_pos*		*Blog_tone_neg*		*Blog_tone_extreme*	
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Competition_con	-0.07	-0.06	0.01	0.01	0.07	0.07	0.04	0.04
	(-1.90)*	(-1.74)*	(0.61)	(0.86)	(2.18)**	(2.10)**	(2.20)**	(2.15)**
Analyst_rec		-0.38		0.01		-0.08		-0.03
		(-5.57)***		(0.55)		(-2.24)**		(-1.72)*
BM		-0.05		-0.06		0.33		0.13
		(-2.27)**		(-2.25)**		(5.04)***		(3.86)***
Size		1.10		-0.00		0.05		0.02
		(5.59)***		(-0.38)		(2.15)**		(1.68)*
Ret		0.38		0.30		-0.80		-0.25
		(6.28)***		(3.90)***		(-4.09)***		(-2.26)**
Momentum		-0.00		0.11		-0.27		-0.08
		(-3.63)***		(4.92)***		(-4.63)***		(-2.41)**
Turnover		0.01		-0.00		0.00		0.00
		(2.45)**		(-1.53)		(3.10)***		(2.18)**
Ananlyst_num		0.08		0.00		-0.01		-0.00
		(2.38)**		(0.27)		(-2.47)**		(-2.07)**
Dispersion		-0.04		-0.03		0.01		-0.01
		(-0.40)		(-0.70)		(0.10)		(-0.20)
Constant	-2.75	-2.25	0.75	0.80	3.50	3.05	2.12	1.92
	(-4.51)***	(-3.31)***	(1.25)	(1.30)	(9.53)***	(6.03)***	(5.44)***	(4.29)***
Blogger Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	47,660	47,660	47,660	47,660	47,660	47,660	47,660	47,660
R-squared	0.46	0.47	0.38	0.38	0.51	0.51	0.50	0.50

Table 6 Analysis Based on Blogger Shutdowns

This table presents results for the following difference-in-difference(s) regression model with blogger, firm, and year-month fixed effects, and with standard errors clustered at the firm level:

$$Blog\_tone_{i,j,t+1} = \beta_0 + \beta_1 \times Treat_j + \beta_2 \times Post_{j,t} + \beta_3 \times Treat_j \times Post_{j,t} + C \times M_{i,j,t} + \varepsilon_{i,t+1},$$

where $Blog\_tone_{i,j,t+1}$ is the average tone of blogs written by blogger j covering stock i in month $t+1$, defined alternatively as the signed difference between the positive tone and the negative tone of blogs (*Blog_tone_net*), the positive tone of blogs (*Blog_tone_pos*), the negative tone of blogs (*Blog_tone_neg*). To form our difference-in-difference(s) sample, we first identify bloggers that exit during our sample period. Treatment firms are those that have been covered by a shutdown blogger during each of the three months prior to the termination date. We employ the propensity score matching (PSM) procedure. In the first stage of the PSM procedure, we model the probability of each firm experiencing a covered blogger's shutdown during this month as a function of following firm characteristics: *Analyst_rec*, *BM*, *Size*, *Ret*, *Momentum*, *Turnover*, *Analyst_num*, and *Dispersion*. We then match with each treatment firm control firms on the condition that the difference between the two firms' predicted probabilities is lower than or equal to 0.01 (the caliber width). $Treat_j$ is an indicator that equals one for our defined treatment firms and zero for matched control firms. $Post_{j,t}$ is an indicator for post-event observations and is zero otherwise. In addition, $M_{i,j,t}$ stacks a set of control variables for stock *i*, and fixed effects for stock *i*, blogger *j* and year-month *t*+1. Appendix A provides definition for all variables. The superscripts ***, **, and * refer to the 1%, 5%, and 10% levels of statistical significance, respectively. The sample includes firm-month observations over the 2006-2011 period.

Analyses Based on Blogger Shutdowns

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	*Blog_tone_net*	*Blog_tone_pos*	*Blog_tone_neg*	*Blog_tone_extreme*	*Blog_tone_net*	*Blog_tone_pos*	*Blog_tone_neg*	*Blog_tone_extreme*
*TREAT*POST*	0.5675	0.0655	-0.5020	-0.2183	0.6652	0.1671	-0.4982	-0.1656
	(2.73)***	(0.29)	(-2.54)**	(-1.17)	(3.39)***	(0.97)	(-2.38)**	(-1.01)
TREAT	-0.5903	-0.0072	0.5831	0.2880	-0.6910	-0.1133	0.5777	0.2322
	(-1.94)*	(-0.05)	(2.21)**	(1.87)*	(-2.16)**	(-0.77)	(1.55)	(0.99)
POST	-0.1118	0.0012	0.1130	0.0571	-0.1143	-0.0084	0.1059	0.0488
	(-1.46)	(0.02)	(1.49)	(1.11)	(-1.49)	(-0.19)	(1.24)	(0.87)
AGE	0.2845	0.4719	0.1874	0.3296	0.3124	0.4780	0.1656	0.3218
	(1.15)	(1.24)	(0.95)	(1.19)	(1.35)	(1.19)	(0.78)	(1.07)
BM	-0.7738	-0.0651	0.7087	0.3218	-1.5651	-0.5280	1.0371	0.2545
	(-3.42)***	(-0.85)	(3.69)***	(3.46)***	(-3.10)***	(-3.00)***	(2.28)**	(1.08)
SIZE	-0.1774	-0.0734	0.1040	0.0153	-0.7416	0.0439	0.7855	0.4147
	(-1.39)	(-1.17)	(0.91)	(0.23)	(-1.79)*	(0.20)	(1.97)*	(1.68)*
Ret	0.0305	0.6520	0.6215	0.6367	0.2377	0.1764	-0.0613	0.0575
	(0.06)	(2.81)***	(1.20)	(2.07)**	(0.55)	(0.65)	(-0.15)	(0.21)
Momentum	0.1228	0.1431	0.0203	0.0817	0.2022	-0.0347	-0.2369	-0.1358
	(0.67)	(2.05)**	(0.13)	(0.98)	(1.13)	(-0.30)	(-1.25)	(-1.05)
Turnover	-0.0072	-0.0021	0.0051	0.0015	-0.0085	-0.0047	0.0038	-0.0004
	(-1.79)*	(-1.04)	(1.44)	(0.74)	(-2.10)**	(-2.01)**	(0.93)	(-0.17)
Analyst_num	0.0074	0.0038	-0.0036	0.0001	-0.0171	0.0194	0.0364	0.0279
	(0.90)	(0.88)	(-0.48)	(0.02)	(-0.80)	(1.76)*	(2.18)**	(3.00)***
Analyst_rec	0.3776	0.0219	-0.3557	-0.1669	0.3748	-0.0637	-0.4386	-0.2512
	(3.85)***	(0.40)	(-4.22)***	(-3.26)***	(2.11)**	(-0.79)	(-2.47)**	(-2.38)**
Dispersion	0.2641	-0.3974	-0.6616	-0.5295	0.9451	-0.3600	-1.3051	-0.8326
	(0.79)	(-2.33)**	(-1.78)*	(-2.24)**	(1.82)*	(-2.16)**	(-2.45)**	(-2.81)***
Constant	-5.1490	9.6065	14.7556	12.1810	6.1312	8.7045	2.5733	5.6389
	(-0.71)	(1.74)*	(1.90)*	(2.16)**	(0.63)	(1.33)	(0.26)	(0.83)
Blogger Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm Fixed Effects	No	No	No	No	Yes	Yes	Yes	Yes
Observations	11,572	11,572	11,572	11,572	11,572	11,572	11,572	11,572
R-squared	0.56	0.37	0.57	0.53	0.58	0.39	0.59	0.54

Table 7 Competition Among Bloggers in Subsamples

This table presents the results for the following regression model with blogger and time fixed effects, and with standard errors clustered at the firm level for subsamples based on analyst coverage, governance quality, and SP500 affiliation:

$$Blog\_tone_{i,j,t+1} = \beta_0 + \beta_1 \times Competition_{i,j,t} + C \times M_{i,j,t} + \varepsilon_{i,t+1},$$

where $Blog\_tone_{i,j,t+1}$ is the average tone of blogs written by blogger *j* covering stock i in month $t+1$, defined alternatively as the signed difference between the positive tone and the negative tone of blogs (*Blog_tone_net*), the positive tone of blogs (*Blog_tone_pos*), the negative tone of blogs (*Blog_tone_neg*), and the degree to which the blog tone is extreme (*Blog_tone_extreme*). In addition, $M_{i,j,t}$ stacks a list of control variables for stock *i* and fixed effects for blogger *j* and year-month *t*+1. Panel A uses *Competition_dummy*, which takes a value of one if the number of bloggers covering the firm—i.e., the competition that a particular blogger faces—is among the top quartile. Panel B uses the continuous value of competition *Competition_con*, which is computed as the natural logarithm of the number of bloggers covering the firm. Appendix A provides definition for all variables. The superscripts ***, **, and * refer to the 1%, 5%, and 10% levels of statistical significance, respectively. The sample includes firm-month observations over the 2006-2011 period.

Panel A						
	Small Analyst_num	*Large Analyst_num*	*Small Govenance*	*Large Govenance*	*Not in SP500*	*In SP500*
Competition_dummy	-0.06	-0.14	-0.08	-0.13	0.04	-0.15
	(-0.67)	(-2.22)**	(-1.22)	(-1.83)*	(0.34)	(-2.75)***
Analyst_tone	0.07	0.15	0.04	0.12	0.08	0.13
	(1.74)*	(2.50)**	(0.84)	(1.87)*	(2.04)**	(2.50)**
BM	-0.30	-0.56	-0.40	-0.48	-0.26	-0.48
	(-3.89)***	(-4.81)***	(-5.65)***	(-4.26)***	(-3.13)***	(-4.77)***
Size	-0.03	-0.07	-0.06	-0.01	-0.11	-0.04
	(-1.46)	(-1.82)*	(-2.32)**	(-0.39)	(-2.10)**	(-0.79)
Ret	0.86	1.44	1.10	0.83	0.44	0.81
	(2.91)***	(5.89)***	(4.22)***	(2.51)**	(1.93)*	(4.40)***
Momentum	0.30	0.49	0.31	0.44	0.44	0.41
	(3.24)***	(6.05)***	(3.83)***	(4.27)***	(4.65)***	(6.04)***
Turnover	0.00	0.00	-0.01	0.00	0.00	0.00
	(-3.29)***	(-2.33)**	(-3.91)***	(-1.67)*	(-2.88)***	(-2.25)**
Analyst_num			0.01	0.00	0.02	0.00
			(1.03)	(0.72)	(2.15)**	(0.72)
Dispersion	-0.03	0.03	0.15	-0.25	0.20	-0.32
	(-0.30)	(0.16)	(1.33)	(-1.38)	(1.55)	(-2.10)**
Constant	-1.37	-2.93	1.89	0.28	0.79	-0.63
	(-1.14)	(-1.98)**	(3.30)***	(0.43)	(1.08)	(-0.85)
Blogger Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	23,462	24,115	21,723	21,812	15,576	30,527
R-squared	0.49	0.51	0.46	0.52	0.50	0.49

Panel B						
	Small Analyst_num	*Large Analyst_num*	*Small Govenance*	*Large Govenance*	*Not in SP500*	*In SP500*
Competition_con	-0.03	-0.09	0.01	-0.12	-0.01	-0.08
	(-0.71)	(-1.84)*	(0.14)	(-2.34)**	(-0.18)	(-2.20)**
Analyst_tone	-0.30	-0.55	-0.41	-0.48	-0.26	-0.48
	(-3.91)***	(-4.80)***	(-5.80)***	(-4.27)***	(-3.11)***	(-4.77)***
BM	-0.03	-0.06	-0.07	0.00	-0.11	-0.04
	(-1.29)	(-1.66)*	(-2.54)**	(0.08)	(-2.08)**	(-0.77)
Size	0.86	1.44	1.10	0.82	0.44	0.81
	(2.89)***	(5.89)***	(4.22)***	(2.45)**	(1.92)*	(4.41)***
Ret	0.30	0.48	0.31	0.43	0.44	0.41
	(3.29)***	(6.01)***	(3.86)***	(4.18)***	(4.64)***	(5.98)***
Momentum	0.00	0.00	-0.01	0.00	0.00	0.00
	(-3.14)***	(-2.23)**	(-4.06)***	(-1.37)	(-2.75)***	(-2.20)**
Turnover			0.00	0.00	0.02	0.00
			(0.92)	(0.83)	(2.25)**	(0.77)
Analyst_num	0.06	0.15	0.03	0.12	0.08	0.13
	(1.71)*	(2.47)**	-0.83	(1.77)*	(2.04)**	(2.45)**
Dispersion	-0.03	0.04	0.15	-0.25	0.20	-0.32
	(-0.31)	(0.21)	(1.29)	(-1.41)	(1.57)	(-2.07)**
Constant	-1.40	-2.98	2.06	-0.05	0.76	-0.57
	(-1.16)	(-2.01)**	(3.52)***	(-0.07)	(1.05)	(-0.78)
Blogger Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	23,462	24,115	21,723	21,812	15,576	30,527
R-squared	0.49	0.51	0.46	0.52	0.50	0.49

Table 8 Impact of Fitted Tone on DGTW-Adjusted Returns

This table presents the results for the following regression model with blogger and time fixed effects and with standard errors clustered at the firm level:

$$DGTW\_ret_{i,t+1} = \beta_0 + \beta_1 \times Fitted\_blog\_tone_{i,t} + \beta_2 \times Residual\_blog\_tone_{i,t} + C \times M_{i,t} + \varepsilon_{i,t+1}.$$

We decompose blog tone into the part due to competition (*Fitted_blog_tone*) and the part unrelated to competition (*Residual_blog_tone*). The decomposition is based on the model: $Blog\_tone_{i,j,t+1} = \beta_0 + \beta_1 \times Competition_{i,j,t} + C \times M_{i,j,t} + \varepsilon_{i,t+1}$. As the first stage is at the blogger firm-month level, we first solve out the fitted value of blog tone at the blogger firm-month level; then, if more than one blogger covered the firm in a month, we aggregate the fitted blog tone to the firm-month level and calculate the residual part of blog tone. Panel A is based on a first stage regression using *Competition_dummy*, which takes a value of one if the number of bloggers covering the firm—i.e., the competition that a particular blogger faces—is among the top quartile. Panel B uses *Competition_con* in the first stage regression, which is computed as the natural logarithm of the number of bloggers covering the firm. $M_{i,t}$ stacks a list of control variables. The superscripts ***, **, and * refer to the 1%, 5%, and 10% levels of statistical significance, respectively. The sample includes 87,442 firm-month observations over the 2006-2011 period.

	Panel A											
	Dependent Variable = *DGTW_ret*											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Fitted_blog_tone_net	0.17		0.22									
	(1.21)		(1.55)									
Residual_blog_tone_net		0.09	0.10									
		(2.19)**	(2.40)**									
News_tone_net	0.07	0.06	0.06									
	(1.27)	(1.12)	(1.06)									
Fitted_blog_tone_pos				0.16		0.16						
				(2.18)**		(2.12)**						
Residual_blog_tone_pos					-0.01	0.01						
					(-0.33)	-0.15						
News_tone_pos				-0.21	-0.21	-0.21						
				(-1.65)	(-1.62)	(-1.65)*						
Fitted_blog_tone_neg							0.04		0.01			
							(0.82)		(0.18)			
Residual_blog_tone_neg								-0.07	-0.07			
								(-2.35)**	(-2.19)**			
News_tone_neg							-0.07	-0.06	-0.06			
							(-1.72)*	(-1.54)	(-1.53)			
Fitted_blog_tone_extreme										0.08		0.06
										(1.36)		(0.92)
Residual_blog_tone_extreme											-0.06	-0.06
											(-1.68)*	(-1.37)
News_tone_extreme										-0.11	-0.10	-0.10
										(-1.77)*	(-1.66)*	(-1.68)*
Analyst_rec	0.27	0.27	0.27	0.27	0.27	0.27	0.27	0.27	0.27	0.27	0.27	0.27
	(3.26)***	(3.29)***	(3.25)***	(3.30)***	(3.29)***	(3.30)***	(3.29)***	(3.29)***	(3.29)***	(3.30)***	(3.29)***	(3.30)***
BM	0.37	0.36	0.37	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36
	(1.42)	(1.40)	(1.41)	(1.40)	(1.40)	(1.40)	(1.41)	(1.40)	(1.40)	(1.40)	(1.40)	(1.40)
Size	-4.29	-4.29	-4.30	-4.30	-4.29	-4.30	-4.29	-4.29	-4.29	-4.29	-4.29	-4.29
	(-20.72)***	(-20.70)***	(-20.76)***	(-20.76)***	(-20.69)***	(-20.77)***	(-20.69)***	(-20.70)***	(-20.70)***	(-20.71)***	(-20.69)***	(-20.71)***
Ret	0.55	0.57	0.54	0.58	0.58	0.58	0.58	0.58	0.58	0.58	0.58	0.58
	(1.19)	(1.25)	(1.17)	(1.26)	(1.27)	(1.26)	(1.27)	(1.26)	(1.26)	(1.27)	(1.27)	(1.27)
Momentum	0.19	0.20	0.18	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20
	(1.35)	(1.43)	(1.32)	(1.43)	(1.44)	(1.43)	(1.45)	(1.43)	(1.44)	(1.45)	(1.44)	(1.45)
Turnover	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02
	(-4.37)***	(-4.49)***	(-4.30)***	(-4.63)***	(-4.52)***	(-4.63)***	(-4.55)***	(-4.48)***	(-4.47)***	(-4.59)***	(-4.49)***	(-4.54)***
Analyst_num	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(-0.21)	(-0.23)	(-0.22)	(-0.28)	(-0.22)	(-0.28)	(-0.24)	(-0.27)	(-0.27)	(-0.25)	(-0.26)	(-0.28)
Dispersion	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17
	(0.53)	(0.52)	(0.52)	(0.54)	(0.53)	(0.54)	(0.53)	(0.52)	(0.53)	(0.54)	(0.53)	(0.53)
Constant	63.19	63.10	63.31	63.05	63.04	63.05	63.00	63.08	63.07	63.00	63.05	63.02
	(20.87)***	(20.79)***	(20.90)***	(20.81)***	(20.80)***	(20.82)***	(20.78)***	(20.80)***	(20.79)***	(20.79)***	(20.79)***	(20.78)***
Firm Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	87,442	87,442	87,442	87,442	87,442	87,442	87,442	87,442	87,442	87,442	87,442	87,442
R-squared	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04

	Panel B											
	Dependent Variable = *DGTW_ret*											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Fitted_blog_tone_net	0.17		0.22									
	(1.26)		(1.60)									
Residual_blog_tone_net		0.09	0.10									
		(2.18)**	(2.40)**									
News_tone_net	0.07	0.06	0.06									
	(1.26)	(1. 12)	(1.06)									
Fitted_blog_tone_pos				0.16		0.16						
				(2.18)**		(2.12)**						
Residual_blog_tone_pos					-0.01	0.01						
					(-0.33)	-0.16						
News_tone_pos				-0.21	-0.21	-0.21						
				(-1.65)	(-1.62)	(-1.65)*						
Fitted_blog_tone_neg							0.04		0.01			
							-0.80		-0.16			
Residual_blog_tone_neg								-0.07	-0.07			
								(-2.34)**	(-2.19)**			
News_tone_neg							-0.07	-0.06	-0.06			
							(-1.71)*	(-1.54)	(-1.53)			
Fitted_blog_tone_extreme										0.08		0.06
										(1.34)		(0.91)
Residual_blog_tone_extreme											-0.06	-0.06
											(-1.67)*	(-1.36)
News_tone_extreme										-0.11	-0.10	-0.10
										(-1.77)*	(-1.66)*	(-1.68)*
Analyst_rec	0.27	0.27	0.27	0.27	0.27	0.27	0.27	0.27	0.27	0.27	0.27	0.27
	(3.26)***	(3.29)***	(3.25)***	(3.30)***	(3.29)***	(3.30)***	(3.29)***	(3.29)***	(3.29)***	(3.30)***	(3.29)***	(3.30)***
BM	0.37	0.36	0.37	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36
	(1.42)	(1. 40)	(1.42)	(1.40)	(1. 40)	(1.40)	(1.41)	(1. 40)	(1.40)	(1.40)	(1. 40)	(1.40)
Size	-4.29	-4.29	-4.30	-4.30	-4.29	-4.30	-4.29	-4.29	-4.29	-4.29	-4.29	-4.29
	(-20.72)***	(-20.70)***	(-20.76)***	(-20.76)***	(-20.69)***	(-20.77)***	(-20.69)***	(-20.70)***	(-20.70)***	(-20.71)***	(-20.68)***	(-20.71)***
Ret	0.55	0.57	0.54	0.58	0.58	0.58	0.58	0.58	0.58	0.58	0.58	0.58
	(1.19)	(1. 25)	(1.17)	(1.26)	(1. 27)	(1.26)	(1.27)	(1. 26)	(1.26)	(1.27)	(1. 27)	(1.27)
Momentum	0.19	0.20	0.18	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20
	(1.34)	(1. 43)	(1.32)	(1.43)	(1. 44)	(1.43)	(1.45)	(1. 43)	(1.44)	(1.45)	(1. 44)	(1.45)
Turnover	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02
	(-4.37)***	(-4.49)***	(-4.30)***	(-4.63)***	(-4.52)***	(-4.63)***	(-4.55)***	(-4.48)***	(-4.47)***	(-4.58)***	(-4.49)***	(-4.54)***
Analyst_num	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(-0.21)	(-0.23)	(-0.22)	(-0.29)	(-0.22)	(-0.28)	(-0.24)	(-0.27)	(-0.27)	(-0.25)	(-0.26)	(-0.28)
Dispersion	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17
	(0.53)	(0. 52)	(0.52)	(0.54)	(0. 53)	(0.54)	(0.53)	(0. 52)	(0.53)	(0.54)	(0. 53)	(0.53)
Constant	63.20	63.10	63.32	63.05	63.04	63.05	63.00	63.08	63.07	63.00	63.05	63.02
	(20.88)***	(20.79)***	(20.91)***	(20.81)***	(20.80)***	(20.82)***	(20.78)***	(20.79)***	(20.79)***	(20.79)***	(20.79)***	(20.78)***
Firm Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Observations	87,442	87,442	87,442	87,442	87,442	87,442	87,442	87,442	87,442	87,442	87,442	87,442
R-squared	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04

Table 9 Competition Among Bloggers With the Peak Year Dummy

This table presents results for the following regression model with blogger and time fixed effects, and with standard errors clustered at the firm level:

$Blog\_tone_{i,j,t+1} = \beta_0 + \beta_1 \times Competition_{i,j,t} + \beta_2 * Competition_{i,j,t} * Peak_t + C \times M_{i,j,t} + \varepsilon_{i,t+1}$,

where $Blog\_tone_{i,j,t+1}$ is average tone of blogs written by blogger j covering stock i in month $t+1$, defined alternatively as the signed difference between the positive tone and the negative tone of blogs (*Blog_tone_net*), the positive tone of blogs (*Blog_tone_pos*), the negative tone of blogs (*Blog_tone_neg*), and the degree to which the blog tone is extreme (*Blog_tone_extreme*). We include the *Peak* dummy to measure a peaking increase in the number of bloggers in 2007 and 2008. $Peak_t$ takes the value of one in the two years of 2007 and 2008 and zero otherwise. Two popular blog platforms emerged in 2007 and 2008. Tumblr was established on Feb 2007, Movable Type, on Dec 2007, and Posterous, on May 2008. The emergence of these blog platforms induced an increase in the number of bloggers to a peak in 2007 and 2008. Panel A uses *Competition_dummy*, which takes a value of one if the number of bloggers covering the firm—i.e., the competition that a particular blogger faces—is among the top quartile. Panel B uses the continuous value of competition, *Competition_con*, which is computed as the natural logarithm of the number of bloggers covering the firm. $M_{i,j,t}$ stacks a list of control variable. Appendix A provides definition for all variables. The superscripts ***, **, and * refer to the 1%, 5%, and 10% levels of statistical significance, respectively. The sample includes 47,660 firm-month observations over the 2006-2011 period.

	Panel A			
	Blog_tone_pos	*Blog_tone_neg*	*Blog_tone_net*	*Blog_tone_extreme*
Competition_dummy	0.02	0.12	-0.10	0.07
	(0.72)	(1.98)**	(-1.60)	(2.06)**
*Competition_dummy*Peak*	-0.09	0.93	-1.02	0.42
	(-0.50)	(2.89)***	(-3.39)***	(1.99)**
Analyst_rec	0.01	-0.08	0.09	-0.04
	-0.52	(-2.31)**	(2.43)**	(-1.79)*
BM	-0.06	0.33	-0.38	0.14
	(-2.24)**	(5.06)***	(-5.59)***	(3.89)***
Size	0.00	0.06	-0.06	0.03
	(-0.25)	(2.63)***	(-2.66)***	(2.15)**
Ret	0.30	-0.81	1.10	-0.25
	(3.89)***	(-4.11)***	(5.61)***	(-2.28)**
Momentum	0.11	-0.27	0.38	-0.08
	(4.91)***	(-4.59)***	(6.23)***	(-2.39)**
Turnover	0.00	0.00	0.00	0.00
	(-1.46)	(3.38)***	(-3.88)***	(2.42)**
Ananlyst_num	0.00	-0.01	0.01	0.00
	(0.30)	(-2.39)**	(2.40)**	(-1.99)**
Dispersion	-0.03	0.00	-0.04	-0.01
	(-0.68)	(0.04)	(-0.32)	(-0.25)
Constant	0.77	2.95	-2.17	1.86
	(1.27)	(5.93)***	(-3.25)***	(4.18)***
Blogger Fixed Effects	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes
Observations	47,660	47,660	47,660	47,660
R-squared	0.38	0.51	0.47	0.50

Panel B				
	Blog_tone_pos	*Blog_tone_neg*	*Blog_tone_net*	*Blog_tone_extreme*
Competition_con	0.01	0.06	-0.05	0.03
	(0.81)	(1.74)*	(-1.40)	(1.80)*
*Competition_con*Peak*	0.01	0.28	-0.27	0.14
	(0.17)	(2.42)**	(-2.09)**	(2.27)**
Analyst_rec	-0.06	0.33	-0.38	0.14
	(-2.25)**	(5.05)***	(-5.58)***	(3.86)***
BM	0.00	0.05	-0.05	0.02
	(-0.38)	(2.23)**	(-2.34)**	(1.75)*
Size	0.30	-0.80	1.10	-0.25
	(3.90)***	(-4.07)***	(5.57)***	(-2.25)**
Ret	0.11	-0.27	0.38	-0.08
	(4.92)***	(-4.62)***	(6.26)***	(-2.40)**
Momentum	0.00	0.00	0.00	0.00
	(-1.53)	(3.13)***	(-3.65)***	(2.20)**
Turnover	0.00	-0.01	0.01	0.00
	(0.28)	(-2.43)**	(2.42)**	(-2.02)**
Ananlyst_num	0.01	-0.08	0.08	-0.03
	-0.55	(-2.21)**	(2.35)**	(-1.69)*
Dispersion	-0.03	0.01	-0.04	-0.01
	(-0.70)	(0.08)	(-0.37)	(-0.23)
Constant	0.79	2.95	-2.16	1.87
	(1.29)	(5.73)***	(-3.16)***	(4.15)***
Blogger Fixed Effects	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes
Observations	47,660	47,660	47,660	47,660
R-squared	0.38	0.51	0.47	0.50

Table 10 Conditional Effects of Blogger Registration and Blogger Location

This table presents the results for the following regression on each blogger of each stock in a monthly period with blogger- and month-fixed effects and with standard errors clustered at the firm level:

$$Blog_{tone_{i,j,t+1}} = \beta_0 + \beta_1 \times Competition_{dummy_{i,j,t}} + \beta_2 \times Competition_{dummy_{i,j,t}} \times Blogger_{char_{i,j,t}} + C \times M_{i,j,t} + \varepsilon_{i,t+1},$$

where $Blog\_tone_{i,j,t+1}$ is the average tone of blogs written by blogger j covering stock i in month $t+1$, defined alternatively as the signed difference between the positive tone and the negative tone of blogs (*Blog_tone_net*), the positive tone of blogs (*Blog_tone_pos*), the negative tone of blogs (*Blog_tone_neg*), and the degree to which the blog tone is extreme (*Blog_tone_extreme*). *Competition_dummy* takes a value of one if the number of bloggers covering the firm—i.e., the competition that a particular blogger faces—is among the top quartile. *Blogger_char* indicates one of the two following variables measuring blogger characteristics: (1) *Register*, an indicator that equals one if a blogger requires user registration and zero otherwise; and (2) *USA*, an indicator that equals one if a blogger's headquarters is located in the USA and zero otherwise. In addition, $M_{i,j,t}$ stacks control variables for stock i and fixed effects for blogger j. Appendix A provides definition for all variables. The superscripts ***, **, and * refer to the 1%, 5%, and 10% levels of statistical significance, respectively. The sample includes firm-month observations over the 2006-2011 period.

	Panel A							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Blog_tone_pos	*Blog_tone_neg*	*Blog_tone_net*	*Blog_tone_extreme*	*Blog_tone_pos*	*Blog_tone_neg*	*Blog_tone_net*	*Blog_tone_extreme*
Competition_dummy	0.08	0.18	-0.10	0.13	0.04	0.13	-0.09	0.09
	(2.47)**	(2.31)**	(-1.30)	(2.91)***	(1.62)	(2.26)**	(-1.51)	(2.61)***
*Competition_dummy*Register*	-0.08	-0.07	-0.01	-0.08				
	(-2.26)**	(-0.86)	(-0.19)	(-1.57)				
*Competition_dummy*USA*					-0.02	0.03	-0.05	0.01
					(-0.55)	(0.45)	(-0.72)	(0.13)
Analyst_rec	0.01	-0.08	0.09	-0.03	0.01	-0.08	0.09	-0.03
	(0.55)	(-2.29)**	(2.43)**	(-1.76)*	(0.54)	(-2.30)**	(2.43)**	(-1.77)*
BM	-0.06	0.33	-0.38	0.13	-0.06	0.32	-0.38	0.13
	(-2.28)**	(5.04)***	(-5.57)***	(3.85)***	(-2.27)**	(5.03)***	(-5.56)***	(3.85)***
Size	-0.01	0.05	-0.06	0.02	-0.00	0.05	-0.06	0.02
	(-0.49)	(2.39)**	(-2.54)**	(1.82)*	(-0.46)	(2.39)**	(-2.53)**	(1.84)*
Ret	0.30	-0.80	1.10	-0.25	0.30	-0.80	1.10	-0.25
	(3.92)***	(-4.11)***	(5.61)***	(-2.26)**	(3.91)***	(-4.11)***	(5.61)***	(-2.27)**
Momentum	0.11	-0.27	0.38	-0.08	0.11	-0.27	0.38	-0.08
	(4.92)***	(-4.65)***	(6.30)***	(-2.42)**	(4.92)***	(-4.65)***	(6.29)***	(-2.41)**
Turnover	-0.00	0.00	-0.00	0.00	-0.00	0.00	-0.00	0.00
	(-1.58)	(3.29)***	(-3.84)***	(2.29)**	(-1.58)	(3.29)***	(-3.84)***	(2.29)**
Analyst_num	0.00	-0.01	0.01	-0.00	0.00	-0.01	0.01	-0.00
	(0.24)	(-2.46)**	(2.44)**	(-2.07)**	(0.23)	(-2.47)**	(2.44)**	(-2.08)**
Dispersion	-0.03	0.01	-0.04	-0.01	-0.03	0.01	-0.04	-0.01
	(-0.69)	(0.12)	(-0.41)	(-0.18)	(-0.69)	(0.12)	(-0.41)	(-0.19)
Constant	1.65	1.64	0.01	1.65	1.65	1.64	0.01	1.65
	(9.47)***	(4.81)***	(0.04)	(8.04)***	(9.43)***	(4.81)***	(0.03)	(8.03)***
Blogger Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	47,660	47,660	47,660	47,660	47,660	47,660	47,660	47,660
R-squared	0.38	0.51	0.47	0.50	0.38	0.51	0.47	0.50

	Panel B							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Blog_tone_pos	*Blog_tone_neg*	*Blog_tone_net*	*Blog_tone_extreme*	*Blog_tone_pos*	*Blog_tone_neg*	*Blog_tone_net*	*Blog_tone_extreme*
Competition	0.01	0.06	-0.05	0.04	0.03	0.09	-0.06	0.06
	(0.92)	(1.70)*	(-1.32)	(1.83)*	(1.70)*	(1.84)*	(-1.17)	(2.21)**
*Competition_con*Register*	-0.01	0.02	-0.03	0.01				
	(-0.32)	(0.62)	(-0.78)	(0.35)				
*Competition_con*USA*					-0.04	-0.03	-0.01	-0.03
					(-1.87)*	(-0.75)	(-0.12)	(-1.35)
Analyst_rec	0.01	-0.08	0.09	-0.03	0.01	-0.08	0.09	-0.03
	(0.55)	(-2.25)**	(2.39)**	(-1.72)*	(0.57)	(-2.24)**	(2.38)**	(-1.70)*
BM	-0.06	0.33	-0.38	0.13	-0.06	0.33	-0.38	0.13
	(-2.25)**	(5.04)***	(-5.57)***	(3.86)***	(-2.27)**	(5.04)***	(-5.58)***	(3.86)***
Size	-0.00	0.05	-0.05	0.02	-0.00	0.05	-0.05	0.02
	(-0.38)	(2.15)**	(-2.27)**	(1.68)*	(-0.42)	(2.14)**	(-2.28)**	(1.65)*
Ret	0.30	-0.80	1.10	-0.25	0.30	-0.80	1.10	-0.25
	(3.90)***	(-4.09)***	(5.59)***	(-2.26)**	(3.91)***	(-4.09)***	(5.59)***	(-2.25)**
Momentum	0.11	-0.27	0.38	-0.08	0.11	-0.27	0.38	-0.08
	(4.92)***	(-4.63)***	(6.27)***	(-2.41)**	(4.92)***	(-4.63)***	(6.28)***	(-2.41)**
Turnover	-0.00	0.00	-0.00	0.00	-0.00	0.00	-0.00	0.00
	(-1.54)	(3.10)***	(-3.63)***	(2.18)**	(-1.55)	(3.09)***	(-3.62)***	(2.16)**
Analyst_num	0.00	-0.01	0.01	-0.00	0.00	-0.01	0.01	-0.00
	(0.27)	(-2.47)**	(2.46)**	(-2.07)**	(0.28)	(-2.47)**	(2.45)**	(-2.06)**
Dispersion	-0.03	0.01	-0.04	-0.01	-0.03	0.01	-0.04	-0.01
	(-0.69)	(0.10)	(-0.39)	(-0.21)	(-0.68)	(0.11)	(-0.40)	(-0.19)
Constant	1.63	1.62	0.01	1.63	1.64	1.62	0.02	1.63
	(9.33)***	(4.61)***	(0.03)	(7.76)***	(9.37)***	(4.61)***	(0.04)	(7.77)***
Blogger Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	47,660	47,660	47,660	47,660	47,660	47,660	47,660	47,660
R-squared	0.38	0.51	0.47	0.50	0.38	0.51	0.47	0.50