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

Yi Dong, Massimo Massa, and Hong Zhang[†]

Abstract

The rise of social media encourages guru dreams due to its low entry barrier and highly skewed distribution of public attention. The pursuing of guru status, however, may be achieved through information provision or cheap talks, and competition inherited in social media may incentivize participants to either process better information or to express options in a more extreme way. Based on a unique dataset of blogs covering S&P 1500 stocks over the period 2006-2011, we find evidence that social media can be informative about future stock return, whereas competition distorts opinions rather than ensuing better information. In particular, competition induces exaggerated negative tones which are unrelated to information. Our results suggest that social media may provide mixed incentives for its participants in terms of information efficiency.

Keywords: Blogs, Social media, Information provision, Competition.
JEL Codes: G30, M41

[†]Yi Dong is with University of International Business and Economics. Massimo Massa is at INSEAD. Hong Zhang is with INSEAD and PBC School of Finance, Tsinghua University. For correspondence: Yi Dong (yidong_ubie@163.com), Massimo Massa (massimo.massa@insead.edu), and Hong Zhang (hong.zhang@insead.edu).

Introduction

One of the most interesting phenomena of the last decade has been the rise and popularity of social media, which differs from traditional media in terms of lower entry barrier and potentially high public attention. Indeed, one of the most important spirits of the new internet era is that it provides opportunities, in an affordable manner, for literally everyone who can use web-based technology to express their opinions. For instance, any individual can, at almost zero cost, create a blog that allows him to express his opinions on almost anything, ranging from stocks, political issues, to fashion, culture and any other topics. More interestingly, given the fast growing body of internet users, blogs or bloggers may attract vast followers to read their opinions—for free. Blogging therefore allows individuals to become salient to attract public attention in a way that is unachievable based on traditional media. The distribution of public attention, however, is likely to be highly skewed, with vast reputation concentrated in a very small fraction of bloggers. These features suggest that we can think of incentives of bloggers as the "dream to become a guru" (e.g., Rosen 1981).

Two interesting questions arise. First, should participants of social media in general and bloggers in particular be somehow informed above and beyond what is covered in the public media, if they want to be a gurus someday? Second, since competition is inherited in social media due to its low entry cost and guru dreams, how could it affect the behavior of bloggers? Potential answers to these questions are crucial in understanding the economics of social media. In this paper, therefore, we collect a unique database of blogs covering all S&P 1500 stocks over the period 2006-2011 from LexisNexis to formally explore these important issues.

We can first compare two alternative hypotheses to explore the potential answer to the first question. On one hand, if the guru status is what bloggers want, we expect them to release some non-public information from time to time in order to build long-term reputation. We can label this hypothesis "*informed guru hypothesis*", in which bloggers are more informed than the public either because they are better able to process information or because they are privy with more private information. The alternative hypothesis (the *cheap-talk hypothesis*) posits that bloggers are not informed above and beyond the public media. Rather, they simply selectively rephrase what is already published in the public media to attract attention.

With respect to the second question, traditional theory has come up with two main theories regarding the effect of competition on information generation. The first theory (the standard competition theory) posits that competition increases the accuracy and reduces potential biases (e.g., Gentzkow and Shapiro, 2006). The second theory (catering theory) posits that the "information producers" structure their report to conform to the views the "information consumers" want to hear.

More competition increases the tendency to cater and therefore distorts the information, reducing its precision and increasing its bias (e.g., Mullainathan and Shleifer, 2005). These two alternative theories share two common features: first, there is a cost for the information provider related to lower accuracy. In the case of the competition theory, this leads to higher accuracy, in the case of the catering theory, this is traded off against the higher benefits of catering. Second, the providers of information in the traditional sense—analysts, brokers, media—all produce information to sell it. Both the cost of inaccuracy and the monetary compensation of information supply, however, are low, if not nonexistent, for bloggers since their incentives are more linked to the Holy Grail of the guru status than direct profitability. These different features suggest that the formulation of the information provision could be even more complicated for bloggers.

We argue that the low entry cost and the highly skewed distribution of gurus intertwine to produce an effective convex benefit for bloggers not only to dig out something big but also to express highly extreme opinions. This intuition is not dissimilar to the traditional wisdom that a convex payoff function encourages risk taking as a response to competition, except that bloggers take additional risk by using more extreme tones to express the same opinion. But what is the best form of extremism to attract public attention? Consolidated psychology literature (e.g., Skowronski and Carlston,1989; Vaish, Grossmann and Woodward, 2008) agrees that negative information tends to influence evaluations more strongly than positive information of similar degree. For instance, as Vaish, Grossmann and Woodward (2008) has pointed out, "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." Hence, bloggers will have an incentive to emphasize negative tones in a more extreme manner when they want to win over the attention-war against their competing peers. Competition, in this particular case, will inflate extreme and especially negative opinions among the competing peers. Resorting to more extreme and negative tones as a response to competition complements the catering theory, and becomes our second working hypothesis (*information distortion hypothesis*). Of course, competition may still affect tones when it incentivizes bloggers to produce more precise information following the standard competition theory, which we label as the *information enhancement alternative hypothesis*.

We test these hypotheses focusing on blogs covering the complete set of S&P 1500 stocks over the period 2006-2011. For each blog we define its tones – *positive tone*, *negative tone* (we remove the negative sign here, hence more negative tone is expressed in a larger numerical value) and *tone difference* in which we net out the score of negative tone from that of positive tone – as well as the degree of *extremism* computed as the maximum value of positive and negative tones of a same blog article. These variables provide a solid cornerstone for us to explore the informativeness of blogs as well as the impact of competition. Of course, for all our tests, we control for similar variables of traditional public media—articles reported by the four largest newspapers in the U.S.

We start by investigating whether bloggers are informed. We first indirectly tests whether the market perceives them to be informed. We therefore relate two stock characteristics that proxy for informed trading and liquidity trading to the presence of blogs. The first proxy is the C2 measure from Llorente, Michaely, Saar, and Wang (2002), which measure the impact of trading volume on return autocorrelation. Since the price impact of liquidity trading tends to reverse, whereas informed trading typically predicts future return, a positive C2 implies informed trading, whereas negative C2 implies liquidity trading. The second proxy is the unformed flows of mutual funds. We document that *blog coverage* is positively related to informed trading and negatively related to uninformed liquidity trading. These results shed initial light on the possibility that blog coverage correlates with informed rather than uninformed trading.

Then, we directly test for informativeness by focusing on the tone of the blog. We find that blog *tone difference* helps to predict abnormal stock performance over the following month. A one-standard-deviation increase in blog tone difference is related to 3.3% higher annualized out-of-sample DGTW return, where the construction of abnormal stock performance follows Daniel et al., (1997). Furthermore, positive tone and negative tone predicts positive and negative DGTW return, respectively. Extremism, by contrast, does not predict future return. Note that blog tones exhibit return predictability even after we explicitly control for the corresponding tones or tone difference of newspaper as well as analyst recommendations, suggesting that bloggers do disseminate information above and beyond what public media provides. Our tests, therefore, provide evidence in favor of the *informed guru hypothesis* rather than the *cheap-talk hypothesis*.

Next we move on to explore the impact of competition in social media. To achieve this goal, we proxy competition by a dummy variable 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 in the cross section, and find a strong negative correlation between competition and *tone difference*. More specifically, competition dummy moves blog *tone difference* from its negative mean further toward the negative direction by another 15%. This impact is both statistically significant and economically sizable. If we further break the analysis into positive and negative tone, we find that there is no significant impact of competition on positive tone, whereas competition significantly enhances the magnitude of negative tone—competition also increases the *extremism* of the tone due to its impact on negative tones.

We also verify that using proxies of competition with continuous values, i.e. when we define competition as the logarithm of the number of competitors, all the above patterns remain robust. In addition, we consider an exogenous event: the change in the number of blogger platforms. Over our sample period, three new popular blogger platforms started in the peak years of 2007-2008—i.e., Tumblr on Feb, 2007, Movable Type on Dec, 2007, and Posterous on May, 2008—and become more stabilized afterwards. Hence we want to specifically examine the impact of competition particularly in these peak years. We find that the increase in the potential competition amplifies the impact of competition in the case of negative tone as well as difference in tone and makes the tone more extreme. During the peak time, for instance, the impact of competition dummy more than doubles the average tone difference.

But does the impact of competition on negative tones come from the processing of more precise negative information or the exaggeration of more negative opinions? We answer this question in two steps. In the first step, we investigate whether such an effect is stronger in the presence of more public information or public scrutiny. More information processing is likely to happen among stocks with less public information or attention, whereas the exaggeration incentives should be stronger especially for stocks are more under public scrutiny because it might be difficult to supply additional information for these stocks. Based on three proxies of public scrutiny, including affiliation with the S&P500 index, analyst coverage and quality of governance, we show that competition among bloggers affects the tone of the blog mostly in firms with high public attention or scrutiny (i.e., high coverage of analysts, better governance and part of the S&P500). This pattern provides initial evidence that competition could exacerbate negative tone rather than encourage more information discovery.

To further confirm that the impact of competition is unrelated to information processing, in our second step of analysis we decompose the difference in tones in blogs into the part induced by competition and the part unrelated to competition (i.e., the rest). We find that the blog tone driven by competition does not have any predictive power in terms of future stock return. In contrast, the part unlinked to competition still exhibits significant predicting power regarding future return, both in terms of tone difference and in terms of negative tone. These results suggest competition, far from increasing the informativeness of the blogs, increases their negative bias, which supports the *information distortion hypothesis* as opposed to the *information enhancement alternative hypothesis*.

Our results shed a new light on the literature exploring how competition affects the dissemination of information in the financial market. It is especially interesting to compare our findings with what we know from the literature about 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. Analysts opinions, for instance, are known to exhibit a positive bias due to conflicts of interest (Brown, Foster, and Noreen, 1985, Stickel, 1990, Abarbanell, 1991, Dreman and Berry, 1995, and Chopra, 1998)—and competition provides a solution to reduce bias and enhance price efficiency (Hong and Kacperczyk 2010; Kelly and Ljungqvist 2012). By contrast, the issue of conflict of interest is minimal 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 pursuing guru status, which illustrate a positive role of social media in terms of information provision, competition—an element at the core of the attractiveness and popularity of social medial—appears to distort information and thus weakens the information contribution of social media. The economics of social media, particularly the part related to information provision, therefore, seem to be completely different from what we have learned from the existing financial market. The general and directional negative blog bias induced by competition also differs from the effect of political polarization often observed in public media (e.g., Groseclose and Milo, 2005).

These differences demonstrate the importance of devoting more research to social media, especially given the rising influence of the latter on our everyday life. Our work, in the respect, contributes to the emerging literature on social media. While a big literature examines the impact of public media on the stock market (e.g., Barber and Loeffler 1993; Huberman and Regev 2001; Busse and Green 2002; 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 is still under-explored. Among the few existing studies, 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) are shown to help information dissemination in the market. Till now, however, blogs – a hugely important social phenomenon – has been ignored in the context of finance. We contribute by showing how blogs are informed and 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 predictability of stock returns. Even more importantly, blogs allow us to explore the impact of competition on social media. Our results shed new light on how competition affects different sectors of the economic differently depending on the incentive structure of participants.

We articulate the rest of the paper as follows. In Section II, we describe the data and the main variables 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 the tone of the blog due to competition. A brief conclusion follows.

II. Data and Main Variables

We collect blog information for all the S&P 1500 stocks from the period from 2006 to 2011. More specifically, the LexisNexis database provides information about the identity of bloggers, the complete text of each blog published by the blogger, the date and time for a blog to be posted, and the keywords of the blog. We retrieve from this data all blogs whose keywords contain any of the S&P 1500 stocks. Appendix 2 provides an example of a blog. We then apply linguistic analyses to each blog in the sample, and link our linguistic analyses to other variables of the firm that we can construct from the CRSP/COMPUSTAT database. In addition to these databases, we obtain analyst information from I/B/E/S, and newspaper articles published on 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 constructed from the above sources. 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 newspaper, the coverage of blogs increases very fast over our sample period from 2006 to 2011, consistent with the gradual popularity of social public network over such 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 largely stays constant, the number of blog articles grows explosively from a mere 3304 in 2006 to 233,040 in 2011. These numbers lay out 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, whose importance is reported in the next a few columns—either in terms of ranks or in terms of market shares.[1] We can see that before 2006, two very big platforms—“Blogger” and “Wordpress”—have already been in operating, though from Panel A we know that the whole size of the blog industry is small. The biggest change in our sample period is in 2007 and 2008, when the two players “Tumblr” and “Posterous” get launched. Since the two players quickly grasp a joint 21% market share as the 2010 poll has illustrated, some exogenous changes have been introduced in terms of the number of blogs competing with each other. Our later tests will use this property to examine the impact of competition.

After the demonstration of the basic industry trend, we next consider the following variables to conduct our empirical tests. The first set of variables is about the tone of blogs. We first process the linguistic content of each blog following Loughran and Mcdonald (2008), which allow us to compute the positive and negative tones of a blog article as weighted value of negative/positive words in the article, denoted as $Blog\_tone\_pos_{i,k,t}$ and $Blog\_tone\_neg_{i,k,t}$, for each blog article k covering stock

[1] 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 poll in different years because no single source provides polls in each year.

i in month t. Lager values in these two variables indicate more positive and more negative tone, respective. In case a blogger posts more than one blog articles for a same firm during the same month, we take the average value of these tone variables. To rule out irrelevant articles that only mention 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 have none-zero scores for both positive and negative tones. To capture the net effect, we also compute the difference between positive and negative tones for each article, denoted as $Blog\_tone\_diff_{i,k,t}$. If blogs are informative, the tone difference should be of special interest. Finally, to capture the degree that an article can use both very positive and very negative words to catch attention, we define extreme tone of a blog article, $Blog\_tone\_extreme_{i,k,t}$, as the max value out of the magnitude of positive tone and that of negative tone, i.e., max($Blog\_tone\_pos_{i,k,t}$, $Blog\_tone\_neg_{i,k,t}$).

For stock-level analyses, we can aggregate blogs at the stock level by averaging over all relevant blogs that cover the same stock on a monthly basis. This steps leads to a set of blog variables, $Blog\_tone\_pos_{i,t}$, $Blog\_tone\_neg_{i,t}$, $Blog\_tone\_diff_{i,t}$, and $Blog\_tone\_extreme_{i,t}$, to refer the average values of positive tones, negative tones, tone difference, and extreme tones of all the blogs covering the same stock in a given month. We can also define blog coverage directly at the firm level as the number of blog articles posted about a firm in a given month.

To explore the impact of competition, we also aggregate blogs at the blogger-stock level by averaging over all relevant blogs written by a same blogger covering a same stock on a monthly basis. This steps leads to $Blog\_tone\_pos_{i,j,t}$, $Blog\_tone\_neg_{i,j,t}$, $Blog\_tone\_diff_{i,j,t}$, and $Blog\_tone\_extreme_{i,j,t}$, which refer the average values of positive tones, negative tones, tone difference, and extreme tones of 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 $News\_tone\_pos_{i,t}$, $News\_tone\_neg_{i,t}$, $News\_tone\_diff_{i,t}$ 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 directly at the firm level as the number of newspaper articles published about a firm in a given month.

We include a set of firm specific dependent or control variables. The *C2* variable comes from Llorente, Michaely, Saar, and Wang (2002), which measures the impact of trading volume on return autocorrelation. The variable *Flow* measures unexpected stock level mutual fund flow based on

Frazzini and Lamont (2008). *DGTW_ret* is the abnormal return following Daniel et al., (1997), in which we adjust stock return by benchmark return constructed from the portfolios that are matched with the stocks held in the evaluated portfolio based on the size, book-to-market and prior-period return characteristics of such stocks.[2]

Among the control variables, *BM* is the book to market ratio. *Size* is the log value of firm total asset. *Ret* is the monthly return. *Momentum* is previous 12 month cumulative return. *Turnover* is monthly volume turnover. *Analyst_num* refers to analyst coverage, calculated as the total number of analyst covered the firm. *Analyst_rec* refers to analyst recommendations, with a larger value referring to a better recommendation (i.e., we reverse the original numerical value of analyst recommendation reported in I/B/E/S, and use 6 minus the median recommendation in the month). Finally, *Dispersion* is the standard deviation of analyst earnings forecast (i.e., EPS) standardized by median analyst earnings forecast. All the variable definitions have been described in appendix A.

We report descriptive statistics on the characteristics of the blog and newspaper coverage in Table 2. In Panel A, we report the summary statistics of stock-level blog and newspaper tone variables, including their whole sample mean, median, standard deviation, and the quintile values at the 25% and 75% of the distribution. We see that newspapers are mildly more extreme than blogs, but the difference between blogs and newspapers is not statistically significant. Panel B, reports the summary statistics of the same list of blog and newspaper variables in the subsample when blog or newspaper coverage is not 0.

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 major variables is reported in Panel D. We can see that blog tone difference positively correlates with DGTW return, and that the magnitude of negative blog tone is especially (negatively) correlated with DGTW return. These observations suggest that blogs could contain useful information about stock return. Of course, whether this is the case or not needs to be tested in multivariate specification, which is the task we will take on next.

III. Are Bloggers Informed?

We recall that our first question asks whether bloggers are informed or whether they simply rely on cheap talk to attract attention. We explore this question in two steps. First, we ask whether the market perceive bloggers to be informed and then we directly test whether they have information.

[2] Detailed description can be found at http://www.rhsmith.umd.edu/faculty/rwermers/ftpsite/DGTW/coverpage.htm.

We start by asking whether the market perceives them to be informed. We expect that if blog coverage is informative, its presence will proxy for the presence of more informed traders and therefore less liquidity traders. We therefore relate some stock characteristics that proxy for informed trading and liquidity trading to the presence of blog coverage in 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 the 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 cluster standard errors at the firm level. (Unreported) results indicate that our results are in general robust to Fama-Macbech specificationss.

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

Next, we directly test for the informativeness of the blogs by focusing on their content "the tone" by 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 tones, including the signed difference between the positive tone and the negative tone of blogs (*Blog_tone_diff*), the positive tone of blogs (*Blog_tone_pos*), the negative tone of blogs (*Blog_tone_neg*), and the degree by which tone is extreme (*Blog_tone_extreme*), and $M_{i,t}$ stacks a list of control variables including newspaper tones, *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 and time level. Note cant, to conduct this test, blog tones are already aggregated at the stock level in a given month.

We report the results in Table 4. We control for analyst recommendations in each model and, to highlight the extent to which blogs can provide information above and beyond public media, alternately tabulate the impact of blog tones without and with the control of similar newspaper tones. The results show the difference in between the positive and negative tones of blogs is highly informative. This holds whether we consider the base specification (Model 2) or whether we control for the degree by which tone is extreme (Model 8). The effect is not only statistically significant, but also economically relevant. One standard deviation higher degree of positive difference in tone is related to 3.3% higher DGTW return.[3]

If we decompose the difference in positive and negative tone, we see that the impact of positive tone is positive while that of negative tone is negative. Note that these predicting powers remain significant even after we control for analyst recommendations and newspaper tones. Hence, both positive and negative tones of blog articles are in general more informative than public media. In contrast, extremism does not seem to have any predicting power on stock return. Note that the presence of newspaper tones typically affects neither the economic magnitude nor the statistical significance of the return predictability of blogs, suggesting that blogs consist of information very different from what the public media provides. Overall, these results support the *informed guru hypothesis*, showing that blogs tend to be on average informed rather than focusing on cheap talks.

IV. Competition and Blog Tone

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

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

where $Blog\_tone_{i,j,t+1}$ is average tone of blog articles 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_diff*), the positive tone of blogs (*Blog_tone_pos*), the negative tone of blogs (*Blog_tone_neg*), and the degree by which tone is extreme (*Blog_tone_extreme*), and $M_{i,j,t}$ stacks control variables for stock i and fixed effect 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

[3] In Model 1, we first compute monthly return impact 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%.

covering the firm—i.e., the competitor that a particular blogger faces—is among the top quartile. 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, in columns (1)-(3), we report the results for the difference in tone, while in columns (4)-(6), we consider positive tone, in columns (7)-(9) negative tone and in columns (10)-(12), we consider the extreme tone.

We see that competition has a very significant and sophisticated impact on the way blog articles are written. In Panel A, Models 1 to 3 demonstrate that the competition dummy typically moves blog *tone difference* further toward the negative direction, and the economic magnitude of the impact is about 15% of its mean value.[4] Consistent with this negative impact, Models (7) to (9) exhibit a clear amplification impact of competition on negative tones. The last three models also show that competition increases the extremism of the tone accordingly. By contrast, and interestingly, competition does not seem to affect positive tones. Panel B further confirm that the impact is robust when we use the continuous proxy for competition. These results provide preliminary evidence in favor of *information distortion hypothesis*, showing that the tone becomes more negatively biased and extreme when competition is higher.

It is interesting to note that the tone of the analysts is instead on average positively related to the tone of the blogs overall. If we break down the tone of the blogs in positive and negative, we see that the analyst tone is negatively related to the negative blog tone and not related to the positive one. This suggests that the blog tone is very different from the one of professional market watchers as the analysts and, in fact, a positive analyst tone is more likely to be negatively related to a negative blog one. Also, the explanatory power of the regression is very high, suggesting that we are indeed pinning down the main determinants of the tones of the blogs.

We also consider an exogenous event: the change in the number of blogger platform. Two popular blogger platforms started on year 2007 and year 2008. This induced a peak increase in blogger number in 2007 and 2008. Tumblr is set up on Feb, 2007, Movable Type is set up on Dec, 2007, and Posterous on May, 2008.Therefore, we estimate:

$$Blog\_tone_{i,t+1} = \beta_0 + \beta_1 \times Competition_{it} + \beta_2 \times Competition_{it} \times Peak_t + C \times M_{i,j,t} + \varepsilon_{i,t+1} \quad (4),$$

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

[4] 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.

We report the results in Table 6, Panel A for *Competition_dummy* and Panel B for *Competition_con*. We see that the peak dummy amplifies the impact of competition in the case of negative tone as well as difference in tone. During the peak time, for instance, the impact of competition dummy more than doubles the average tone difference.[5] It is also makes the tone more extreme. In contrast, in line with our expectations, there is no impact on positive tone.

V. Blogs and Information

A further confirmation of the information distortion hypothesis, however, requires us to directly investigate whether blog tone is more negative due to more precise information or a simple exaggeration of more extreme tones without providing any additional information.

We therefore break down the relationship between tone of the blog and competition in different sub-samples defined in terms of *Analyst_num*, the number of analysts covering the firm in a year, governance (Aggarwal et al 2009) and whether the firm is in SP500 index. We report the results in Table 7. We see that competition among bloggers affects the tone of the blog mostly in firms with high coverage of analysts, better governance and part of the S&P500. More specifically competition exacerbates the negative tone of the bloggers especially in the case in which the stocks are more under the media scrutiny. These results support the information distortion hypothesis.

Finally, we add the pieces together and we ask whether the link between blog tone and returns is due to the effect of competition among the bloggers. To investigate this issue, we first decompose the degree of blog tone into the part due to competition ("fitted blog tone") and the part unrelated to it ("residual blog tone") and then we relate these two orthogonal components to stock returns. More specifically, we estimate:

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

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 tones induced and unrelated to competition, respectively. We apply this decomposition to all four variables related to blog tones, including *Blog_tone_diff*, *Blog_tone_pos*, *Blog_tone_neg*, and *Blog_tone_extreme*, and report the results in Table 8. In columns (1)-(3), we report the results for the overall tone, in columns (4)-(6), for the positive tone, in columns (7)-(9), for the negative tone and in columns (10)-(12) for the degree of blog extremism in blog tone.

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

We see that the blog tone driven by competition does not have any predictive power in terms of future stock return. In contrast, the residual component – i.e., the part not linked to the distortionary effect of competition – predicts future return, both in terms of tone difference and in terms of negative tone. In particular, one standard deviation higher residual tone difference (negative tone) predicts a 1.1% (1.09%) annualized abnormal return.[6] This confirms that the tone (and especially its negative part) helps to predict returns. Indeed, the tone may be related to information or sentiment in the market. The first can be based on some private information not yet impounded into the prices, while the second can be based on the impact of sentiment. In other words, blogs proxy for the sentiment behavior of the market or "dumb money" (e.g., Frazzini and Lamont, 2008) and therefore proxy for stock overvaluation. In contrast, the part of the tone just related to competition has no predictive power, as it is just distorted in its informational content by the frantic rush of the bloggers to outsmart them. These results confirm the information distortion hypothesis.

Conclusion

In this paper, we study the economics of social media based on a unique dataset of blogs. Compared to traditional media, social media features lower entry barrier and potentially high public attention, which allows participants to pursue guru status based on the articles they posed. This new phenomenon leads to two important questions of whether social media attracts attention via information processing or via cheap talking and whether competition intensifies the incentive of information discovery or distorts the tones of options.

We document that bloggers are informed and are able on average to predict risk-adjusted stock performance, suggesting that social media can supply information above and beyond public media. However, competition in general leads to more exaggerated negative tones with little return predicting power, implying that competition in social media distorts information. This impact differs drastically from what we observe in other parts of the economy—competition for instance improves the accuracy of information supplied by analysts. Our results, therefore, shed a new light not only on the economics of social media but also on how competition affects the dissemination of information in our economy.

[6] Similar to Table 3, we first compute monthly return impact 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 fitted tone difference. We then annualize the compounded impact of 0.092% as 1.1%. Model (9) allows us to compute the monthly return impact 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 fitted tone difference. We then annualize the compounded impact of 0.0903% as 1.09%.

References

Abarbanell, Jeffery S., 1991, Do analysts' earnings forecasts incorporate information in prior stock price changes?, Journal of Accounting and Economics 14, 147-165.

Antweiler,W., and M. Z. Frank. 2004. Is all that talk just noise? The information content of internet stock message

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

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

Brown, Phillip, George Foster, and Eric Noreen, 1985, Security analyst multi-year earnings forecasts and the capital market, American Accounting Association: Sarasota, FL.

Busse, J. A., and T. C. Green. 2002. Market efficiency in real time. Journal of Financial Economics

65:415–37.

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

Chopra,Vijay K., 1998, Why so much error in analysts' earnings forecasts?, Financial Analysts Journal 54, 30-37.

Daniel, K., M. Grinblatt, S. Titman, and R. Wermers, 1997, Measuring mutual fund pPerformance with characteristic-based benchmarks, Journal of Finance 52, 1035-58.

Das, S. R., and M. Y. Chen. 2007. Yahoo! for Amazon: Sentiment extraction from small talk on the web. Management Science 53:1375–88.

Dreman, David and Michael Berry, 1995, Analyst forecasting errors and their implications for security analysis, Financial Analysts Journal 51, 30-42.

Dougal, C., J. Engelberg, D. Garcia, and C. Parsons. 2012. Journalists and the stock market. Review of Financial

Studies 25:639–79.

Engelberg, J. 2008. Costly information processing: Evidence from earnings announcements. SSRN eLibrary.

Engelberg, J., and C. A. Parsons. 2011. The causal impact of media in financial markets. Journal of Finance 66:67–97.

Fang, L., and J. Peress. 2009. Media coverage and the cross-section of stock returns. Journal of Finance 64:2023–52.

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

Gentzkow, Matthew and Jesse M. Shapiro, 2006, Media bias and reputation, Journal of Political Economy 114, 280-316.

Groseclose, Tim and Jeffrey Milyo, 2005, A measure of media bias, Quarterly Journal of Economics 120, 1191–1237.

Gurun, U. G., and A. W. Butler. 2012. Don't believe the hype: Local media slant, local advertising, and firm value. Journal of Finance 67:561–98.

Hong, H. and M. Kacperczyk, 2010. Competition and bias. Quarterly Journal of Economics 125, 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 56:387–96.

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

Mullainathan, Sendhil and Andrei Shleifer, 2005, The market for news, American Economic Review 95, 1031–1053.

Rosen, Sherwin, 1981, The Economics of Superstars, American Economic Review 71, 845–858.

Solomon, D. H. 2012. Selective publicity and stock prices. Journal of Finance 67:599–638.

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

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

Tetlock, P. C., 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. F. Whitelaw. 2001. News or noise? Internet postings and stock prices. Financial Analysts Journal 57:41–51.

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

Appendix A Variable Definitions

Variable Name	Variable Definitions
Blog Related Variables	
Blog_coverage	The number of blog articles covered the firm in a month;
Blog_tone_pos	The average value of postive tone (weighted value of postive word following Loughran and Mcdonald (2008)) of all articles covered the firm in a month;
Blog_tone_neg	The average value of negative tone (weighted value of postive word following Loughran and Mcdonald (2008)) of all articles covered the firm in a month;
Blog_tone_diff	The signed difference between the positive tone and the negative tone of blogs covered the firm in a month;
Blog_tone_extreme	The max value out of the magnitude of positive tone and that of negative tone of blogs covered the firm in a month;
Competition_dummy	It 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	It computed as the logarithm of the number of bloggers covering the firm;
Age	It meansures the age of blogger, which is the number of month from the first time the blogger appeared in the database till current month;
Peak	The dummy variables that takes a value of 1 in the two years of 2007 and 2008, and 0 otherwise.
Newspaper Related Varaibles	
News_coverage	The number of blog articles covered the firm in a month;
News_tone_pos	The average value of postive tone (weighted value of postive word following Loughran and Mcdonald (2008)) of all articles covered the firm in a month;
News_tone_neg	The average value of negative tone (weighted value of postive word following Loughran and Mcdonald (2008)) of all articles covered the firm in a month;
News_tone_diff	The signed difference between the positive tone and the negative tone of blogs covered the firm in a month;
News_tone_extreme	The max value out of the magnitude of positive tone and that of negative tone of blogs covered the firm in a month.
Other Main Variables	
C2	It comes from Llorente, Michaely, Saar, and Wang (2002), which measures the impact of trading volume on return autocorrelation;
DGTW_ret	Abnormal return following Daniel et al., (1997), in whichweadjust stock return by benchmark return constructed from the portfolios that are matched with the stocks held in the evaluated portfolio based on the size, book-to-market and prior-period return characteristics of such stocks;
Flow	It measures unexpected stock level mutual fund flow based on Frazzini and Lamont (2008).
Control Variables	
Analyst_num	It refers to analyst coverage, calculated as the total number of analyst covered the firm;
Analyst_rec	It refers to analyst recommendations, with a larger value referring to a better recommendation;
BM	The book to market ratio;
Dispersion	The standard deviation of analyst earnings forecast (i.e., EPS) standardized by median analyst earnings forecast;
Momentum	Previous 12 month cumulative return;
Ret	Monthly return;
Size	The log value of firm total asset;
Turnover	Monthly volume turnover.

Appendix B Example of Blog Article

Below is an example from LexisNexis, by DCist blog about frim 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 shows the time series summary statistics of blogs and large blog platforms. 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. The final two columns report the number of newspaper articles and the number of blogs in a given year. Panel B reports the launching year for some of the largest blog platforms, whose importance is reported in the next a few columns—either in terms of ranks or in terms of 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 use the different website poll in different years because no single source provides polls in each year. Our sample covers 2006 to 2011.

Panel A

Year	# of firms with blog coverage	# of firms with newspaper coverage	# of bloggers	# of newspaper articles	# of blog articles
2006	653	634	206	7004	3304
2007	1093	639	747	6986	16739
2008	1270	638	1530	6249	34005
2009	1366	599	1882	5276	67177
2010	1428	576	2066	4616	144735
2011	1415	537	2195	3843	233040

Panel B

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 of Main Variables

This table shows the summary statistics of our main and control variables. Panel A reports the summary statistics of blog coverage, blog tone, newspaper coverage and newspaper tone. Panel B report the summary statistics of blog coverage, tone in the conditional sample, when the firm month has been covered by at least one blog articles. And we also report the summary statistics of newspaper coverage and tone when the firm month has been covered by at least one newspaper articles. Panel C shows the summary statistics of other variables in the following regressions. Panel D reports the pearson correlation between other firm month variable in the following regression. All the variable definitions have been described in appendix A.

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_diff	1.41	-0.18	0	0	0
News_tone_diff	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
Blog_tone_conflict	0.23	0.11	0	0	0
News_tone_conflict	0.10	0.02	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					
News_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					
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_rec	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_diff	0.007	-0.013	0.002	-0.21	-0.156	1							
(6)	0.054	<.0001	0.625	<.0001	<.0001								
News_tone_diff	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 Coverage

This table conducts the following regression for each stock in a monthly period with firm and month fixed effect, and cluster at 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 C2 and Flow, for stock i in period $t+1$.C2 comes from Llorente, Michaely, Saar, and Wang (2002), which measures the impact of trading volume on return autocorrelation. Flow measures unexpected stock level mutual fund flow based on 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. All variables have been described in the appendix A. 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 periods.

	Dependent Variable = *C2*			Dependent Variable = *Flow*		
	(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)***
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 conducts the following regression for each stock in a monthly period with firm and month fixed effect, and cluster at firm level,

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

where$DGTW_ret_{i,t+1}$ is the out-of-sample abnormal performance of stock i in month$t+1$, (i.e. abnormal return following Daniel et al., (1997), in which we adjust stock return by bench mark return constructed from the portfolios that are matched with the stocks held in the evaluated portfolio based on the size, book-to-market and prior-period return characteristics of such stocks.) $Blog\_tone_{i,t}$ refers to the list of variables describing blog tones, including the signed difference between the positive tone and the negative tone of blogs (*Blog_tone_diff*), the positive tone of blogs (*Blog_tone_pos*), the negative tone of blogs (*Blog_tone_neg*), and the degree by which tone is extreme (*Blog_tone_extreme*), and $M_{i,t}$ stacks a list of control variables including newspaper tones. All variables have been described in the appendix A. 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 periods.

	Dependent Variable = *DGTW_ret*							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Blog_tone_diff	0.10	0.10					0.11	0.11
	(2.64)***	(2.54)**					(2.67)***	(2.61)***
News_tone_diff		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)***
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 conducts the following regression for each blogger of each stock in a monthly period with blogger and month fixed effect, and cluster at 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_diff*), the positive tone of blogs (*Blog_tone_pos*), the negative tone of blogs (*Blog_tone_neg*), and the degree by which tone is extreme (*Blog_tone_extreme*), and $M_{i,j,t}$ stacks control variables for stock $i$and fixed effect for blogger j. Panel A uses the *Competition_dummy*, 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. Panel B is using continuous value of competition which is computed as the logarithm of the number of bloggers covering the firm. $M_{i,j,t}$ stacks a list of control variables including blogger age and newspaper coverage. Other control variables have been described in the appendix A. 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 periods.

Panel A												
	Dependent Variable = *Blog_tone_diff*			Dependent Variable = *Blog_tone_pos*			Dependent Variable = *Blog_tone_neg*			Dependent Variable = *Blog_tone_extreme*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Competition_dummy	-0.12	-0.11	-0.11	0.03	0.03	0.03	0.15	0.14	0.14	0.09	0.09	0.09
	(-2.01)**	(-2.10)**	(-2.10)**	(1.32)	(1.54)	(1.54)	(2.61)***	(2.85)***	(2.85)***	(2.89)***	(3.08)***	(3.08)***
Age			0.01			0.01			0.00			0.00
			(1.14)			(0.91)			(-0.35)			(0.50)
Analyst_rec		-0.38	-0.38		-0.06	-0.06		0.33	0.33		0.13	0.13
		(-5.57)***	(-5.57)***		(-2.27)**	(-2.27)**		(5.03)***	(5.03)***		(3.85)***	(3.85)***
BM		-0.06	-0.06		0.00	0.00		0.05	0.05		0.02	0.02
		(-2.53)**	(-2.53)**		(-0.46)	(-0.46)		(2.40)**	(2.40)**		(1.84)*	(1.84)*
Size		1.10	1.10		0.30	0.30		-0.80	-0.80		-0.25	-0.25
		(5.61)***	(5.61)***		(3.91)***	(3.91)***		(-4.11)***	(-4.11)***		(-2.27)**	(-2.27)**
Ret		0.38	0.38		0.12	0.12		-0.27	-0.27		-0.08	-0.08
		(6.30)***	(6.30)***		(4.92)***	(4.92)***		(-4.65)***	(-4.65)***		(-2.41)**	(-2.41)**
Momentum		0.00	0.00		0.00	0.00		0.00	0.00		0.00	0.00
		(-3.84)***	(-3.84)***		(-1.58)	(-1.58)		(3.29)***	(3.29)***		(2.29)**	(2.29)**
Turnover		0.01	0.01		0.00	0.00		-0.01	-0.01		0.00	0.00
		(2.43)**	(2.43)**		(0.23)	(0.23)		(-2.46)**	(-2.46)**		(-2.08)**	(-2.08)**
Ananlyst_num		0.09	0.09		0.01	0.01		-0.08	-0.08		-0.03	-0.03
		(2.42)**	(2.42)**		(0.53)	(0.53)		(-2.29)**	(-2.29)**		(-1.77)*	(-1.77)*
Dispersion		-0.04	-0.04		-0.03	-0.03		0.01	0.01		-0.01	-0.01
		(-0.41)	(-0.41)		(-0.70)	(-0.70)		(0.12)	(0.12)		(-0.18)	(-0.18)
Constant	-2.76	-2.22	-0.18	0.75	0.81	1.34	3.50	3.03	1.52	2.13	1.92	1.43
	(-4.51)***	(-3.30)***	(-0.45)	(1.25)	(1.32)	(4.59)***	(9.40)***	(6.06)***	(4.20)***	(5.41)***	(4.30)***	(5.56)***
Observations	47,660	47,660	47,660	47,660	47,660	47,660	47,660	47,660	47,660	47,660	47,660	47,660
R-squared	0.46	0.47	0.47	0.38	0.38	0.38	0.51	0.51	0.51	0.50	0.50	0.50

Panel B												
	Dependent Variable = *Blog_tone_diff*			Dependent Variable = *Blog_tone_pos*			Dependent Variable = *Blog_tone_neg*			Dependent Variable = *Blog_tone_extreme*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Competition_con	-0.07	-0.06	-0.06	0.01	0.01	0.01	0.07	0.07	0.07	0.04	0.04	0.04
	(-1.90)*	(-1.74)*	(-1.74)*	(0.61)	(0.86)	(0.86)	(2.18)**	(2.10)**	(2.10)**	(2.20)**	(2.15)**	(2.15)**
Age			0.01			0.01			-0.00			0.00
			(1.19)			(0.91)			(-0.44)			(0.46)
Analyst_rec		-0.38	-0.38		0.01	0.01		-0.08	-0.08		-0.03	-0.03
		(-5.57)***	(-5.57)***		(0.55)	(0.55)		(-2.24)**	(-2.24)**		(-1.72)*	(-1.72)*
BM		-0.05	-0.05		-0.06	-0.06		0.33	0.33		0.13	0.13
		(-2.27)**	(-2.27)**		(-2.25)**	(-2.25)**		(5.04)***	(5.04)***		(3.86)***	(3.86)***
Size		1.10	1.10		-0.00	-0.00		0.05	0.05		0.02	0.02
		(5.59)***	(5.59)***		(-0.38)	(-0.38)		(2.15)**	(2.15)**		(1.68)*	(1.68)*
Ret		0.38	0.38		0.30	0.30		-0.80	-0.80		-0.25	-0.25
		(6.28)***	(6.28)***		(3.90)***	(3.90)***		(-4.09)***	(-4.09)***		(-2.26)**	(-2.26)**
Momentum		-0.00	-0.00		0.11	0.11		-0.27	-0.27		-0.08	-0.08
		(-3.63)***	(-3.63)***		(4.92)***	(4.92)***		(-4.63)***	(-4.63)***		(-2.41)**	(-2.41)**
Turnover		0.01	0.01		-0.00	-0.00		0.00	0.00		0.00	0.00
		(2.45)**	(2.45)**		(-1.53)	(-1.53)		(3.10)***	(3.10)***		(2.18)**	(2.18)**
Ananlyst_num		0.08	0.08		0.00	0.00		-0.01	-0.01		-0.00	-0.00
		(2.38)**	(2.38)**		(0.27)	(0.27)		(-2.47)**	(-2.47)**		(-2.07)**	(-2.07)**
Dispersion		-0.04	-0.04		-0.03	-0.03		0.01	0.01		-0.01	-0.01
		(-0.40)	(-0.40)		(-0.70)	(-0.70)		(0.10)	(0.10)		(-0.20)	(-0.20)
Constant	-2.75	-2.25	-0.20	0.75	0.80	1.32	3.50	3.05	1.52	2.12	1.92	1.42
	(-4.51)***	(-3.31)***	(-0.48)	(1.25)	(1.30)	(4.52)***	(9.53)***	(6.03)***	(4.10)***	(5.44)***	(4.29)***	(5.45)***
Observations	47,660	47,660	47,660	47,660	47,660	47,660	47,660	47,660	47,660	47,660	47,660	47,660
R-squared	0.46	0.47	0.47	0.38	0.38	0.38	0.51	0.51	0.51	0.50	0.50	0.50

Table 6 Competition among Blogger with Peak Year Dummy

This table conducts the following regression for each blogger of each stock in a monthly period with blogger and month fixed effect, and cluster at 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_diff*), the positive tone of blogs (*Blog_tone_pos*), the negative tone of blogs (Blog_tone_neg), and the degree by which tone is extreme (Blog_tone_extreme). We include Peak_dummy to measure a peak increase in blogger number in 2007 and 2008. We consider an exogenous event: the change in the number of blogger platform. Two popular blogger platforms started on year 2007 and year 2008. This induced a peak increase in blogger number in 2007 and 2008. Tumblr is set up on Feb, 2007, Movable Type is set up on Dec, 2007, and Posterous on May, 2008. $\mathrm{Peak_t}$ is the dummy variables that takes a value of 1 in the two years of 2007 and 2008, and 0 otherwise. Panel A uses the *Competition_dummy*, 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. Panel B is using continuous value of competition which is computed as the logarithm of the number of bloggers covering the firm. $\mathrm{M_{i,j,t}}$ stacks a list of control variables including blogger age and newspaper coverage. Other control variables have been described in the appendix A. 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 periods.

	Panel A			
	Dependent Variable = *Blog_tone_Pos*	Dependent Variable = *Blog_tone_Neg*	Dependent Variable = *Blog_tone_Diff*	Dependent Variable = *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)***
Observations	47,660	47,660	47,660	47,660
R-squared	0.38	0.51	0.47	0.50

Panel B				
	Dependent Variable = *Blog_tone_Pos*	Dependent Variable = *Blog_tone_Neg*	Dependent Variable = *Blog_tone_Diff*	Dependent Variable = *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)***
Observations	47,660	47,660	47,660	47,660
R-squared	0.38	0.51	0.47	0.50

Table 7 Competition among Bloggers in Subsamples

This table conducts the following regression for each blogger of each stock in a monthly period with blogger and month fixed effect, and cluster at firm level in each subsample separated by analyst coverage, corporate governance and whether the firm is in SP500 index,

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

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_diff*), the positive tone of blogs (*Blog_tone_pos*), the negative tone of blogs (*Blog_tone_neg*), and the degree by which tone is extreme (*Blog_tone_extreme*), and $M_{i,j,t}$ stacks control variables for stock $i$and fixed effect for blogger j. Panel A uses the *Competition_dummy*, 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. Panel B is using continuous value of competition which is computed as the logarithm of the number of bloggers covering the firm. $M_{i,j,t}$ stacks a list of control variables including blogger age and newspaper coverage. Other control variables have been described in the appendix A. 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 periods.

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_rec	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)
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_rec	-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)
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 Return

This table conducts the following regression for each blogger of each stock in a monthly period with blogger and month fixed effect, and cluster at 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 the degree of blog tone into the part due to competition ("Fitted blog tone") and the part unrelated to it ("Residual blog tone"),where $Fitted\_blog\_tone_{i,t}$ refers to the fitted blog tone due to competition, and $Residual\_Blog\_Tone_{i,t}$ refers to the residual blog tone which is unrelated to competition. The decomposition is based on model $Blog\_tone_{i,j,t+1} = \beta_0 + \beta_1 \times Competition_{i,j,t} + C \times M_{i,j,t} + \varepsilon_{i,t+1}$, while the first stage is at blogger firm month level, we first solve out the fitted value of blog tone at blogger firm month level, if there is more than one blogger covered the firm in a month, then we aggregate the fitted blog tone into firm month level, and calculated the residual part. Panel A bases on first stage regression of *Competition_dummy*, which 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 panel B bases on *Competition_con*, which is computed as the logarithm of the number of bloggers covering the firm. $M_{i,t}$ stacks a list of control variables including newspaper coverage. Other control variables have been described in the appendix A. 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 periods.

Panel A															
	Dependent Variable = *DGTW_ret*														
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Fitted_blog_tone_diff	0.17		0.22												
	(1.21)		(1.55)												
Residual_blog_tone_diff		0.09	0.10										0.08	0.07	0.07
		(2.19)**	(2.40)**										(2.05)**	(1.74)*	(1.73)*
News_tone_diff	0.07	0.06	0.06										-0.05	-0.05	-0.05
	(1.27)	(1.12)	(1.06)										(-0.49)	(-0.51)	(-0.50)
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	0.06		0.06
										(1.36)		(0.92)	(1.04)		(0.89)
Residual_blog_tone_extreme											-0.06	-0.06		-0.03	-0.02
											(-1.68)*	(-1.37)		(-0.71)	(-0.47)
News_tone_extreme										-0.11	-0.10	-0.10	-0.15	-0.15	-0.15
										(-1.77)*	(-1.66)*	(-1.68)*	(-1.30)	(-1.29)	(-1.30)
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	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)***	(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	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)	(1.39)	(1.39)	(1.39)
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	-4.30	-4.29	-4.30
	(-20.72)***	(-20.70)***	(-20.76)***	(-20.76)***	(-20.69)***	(-20.77)***	(-20.69)***	(-20.70)***	(-20.70)***	(-20.71)***	(-20.69)***	(-20.71)***	(-20.73)***	(-20.72)***	(-20.74)***
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	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)	(1.26)	(1.26)	(1.26)
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	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)	(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	-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)***	(-4.54)***	(-4.48)***	(-4.53)***
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	0.00	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)	(-0.26)	(-0.24)	(-0.27)
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	0.17	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)	(0.53)	(0.52)	(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	63.08	63.11	63.08
	(20.87)***	(20.79)***	(20.90)***	(20.81)***	(20.80)***	(20.82)***	(20.78)***	(20.80)***	(20.79)***	(20.79)***	(20.79)***	(20.78)***	(20.80)***	(20.81)***	(20.80)***
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	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	0.04	0.04	0.04

	Panel B														
	Dependent Variable = *DGTW_ret*														
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Fitted_blog_tone_diff	0.17		0.22												
	(1.26)		(1.60)												
Residual_blog_tone_diff		0.09	0.10										0.08	0.07	0.07
		(2.18)**	(2.40)**										(2.04)**	(1.73)*	(1.72)*
News_tone_diff	0.07	0.06	0.06										-0.05	-0.05	-0.05
	(1.26)	(1.12)	(1.06)										(-0.49)	(-0.51)	(-0.50)
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	0.06		0.05
										(1.34)		(0.91)	(1.03)		(0.88)
Residual_blog_tone_extreme											-0.06	-0.06		-0.03	-0.02
											(-1.67)*	(-1.36)		(-0.70)	(-0.47)
News_tone_extreme										-0.11	-0.10	-0.10	-0.15	-0.15	-0.15
										(-1.77)*	(-1.66)*	(-1.68)*	(-1.30)	(-1.29)	(-1.30)
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	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)***	(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	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)	(1.39)	(1.39)	(1.39)
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	-4.30	-4.29	-4.30
	(-20.72)***	(-20.70)***	(-20.76)***	(-20.76)***	(-20.69)***	(-20.77)***	(-20.69)***	(-20.70)***	(-20.70)***	(-20.71)***	(-20.68)***	(-20.71)***	(-20.73)***	(-20.71)***	(-20.73)***
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	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)	(1.26)	(1.26)	(1.26)
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	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)	(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	-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)***	(-4.54)***	(-4.48)***	(-4.53)***
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	0.00	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)	(-0.26)	(-0.24)	(-0.27)
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	0.17	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)	(0.53)	(0.52)	(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	63.08	63.10	63.08
	(20.88)***	(20.79)***	(20.91)***	(20.81)***	(20.80)***	(20.82)***	(20.78)***	(20.79)***	(20.79)***	(20.79)***	(20.79)***	(20.78)***	(20.80)***	(20.81)***	(20.80)***
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	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	0.04	0.04	0.04