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 the low entry barrier and highly skewed distribution of public attention that characterize social media. 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 express more extreme options. Using a unique dataset of blogs covering S&P 1500 stocks over the 2006-2011 period, we find evidence that social media can be informative about future stock returns but that competition distorts opinions rather than encouraging participants to process better information. In particular, competition induces exaggerated negative tones in blogs, which is 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. Email: yidong_ubie@163.com. 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.

Introduction

One of the most interesting phenomena of the last decade has been the rise in popularity of social media. Unlike traditional media, social media are characterized by a low barrier to entry and very high potential for speedy 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 important, 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 the 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" (e.g., Rosen, 1981).

Two interesting questions arise: First, should "gurus-to-be" bloggers be more informed than public media? Second, given the low entry costs of blogging, how does competition affect bloggers' behavior and shape their dreams to become gurus? The answers to these questions are crucial for understanding the economics of social media.

In this paper, we address these questions by using a unique hand-collected database of blogs covering all S&P 1500 stocks over the 2006-2011 period from LexisNexis. We start by investigating whether bloggers are informed. We entertain two alternative hypotheses. The first posits that if the objective of bloggers is to become "gurus," bloggers must release some non-public information to build a long-term reputation. Bloggers may be more informed than the public either because they are better able to process information or because they are privy to more private information. We label this hypothesis the *informed guru hypothesis*. The alternative hypothesis posits that bloggers are not more informed than public media; rather, they simply selectively rephrase what is already published in the public media to attract attention. We label this hypothesis the *cheap-talk hypothesis*.

We first investigate whether blog coverage is related to more informed trading in general. We therefore relate the presence of blogs to two stock characteristics that proxy for informed trading and liquidity trading. The first proxy is the C2 measure from Llorente, Michaely, Saar, and Wang (2002), which exploits the impact of trading volume on return autocorrelation to distinguish between liquidity trading and informed 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 is correlated with informed rather than uninformed trading.

Next, we directly test for the informativeness of blogs by focusing on the stock return predictability of their tone. For each blog article, we follow Loughran and Mcdonald (2008) and use linguistic analysis to define the tone of a blog, including its *positive tone* (based on the distribution of positive words in the blog), its *negative tone* (based on the distribution of negative words in the blog), and its *tone difference* (computed as the difference between *positive* and *negative tone*). We also define the degree of tone *extremism* as the maximum value for the positive and negative tone of the same blog article.

We find that blog *tone difference* helps to predict abnormal stock performance over the following month. Specifically, a one-standard-deviation increase in blog tone difference is related to a 3.3% higher annualized out-of-sample DGTW stock return, for which the construction of abnormal stock performance follows Daniel et al. (1997). Furthermore, positive tone and negative tone predict positive and negative DGTW returns, respectively. By contrast, extremism does not predict future returns. Importantly, blog tone exhibits return predictability even after we explicitly control for the corresponding tone or tone difference 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 public media provides. Our tests therefore provide evidence in favor of the *informed guru hypothesis* rather than the *cheap-talk hypothesis*.

Once we have established the informativeness of blogs, we move on to explore the impact of competition in the blog market. In this case, theory also provides us with two alternative hypotheses. First, standard competition theory posits that competition increases the accuracy and reduces the potential biases of information (e.g., Gentzkow and Shapiro, 2006). If we apply the same intuition to social media, competition would incentivize bloggers to produce more precise information. We call this prediction the *information enhancement hypothesis*.

However, “information producers” may also have incentives to structure their report to cater to what “information consumers” want to hear. More competition in this case may increase information producers’ tendency to cater and therefore distort information, reducing its precision and increasing its bias (e.g., Mullainathan and Shleifer, 2005). We argue that this alternative effect may also apply to social media, especially because of the two characteristics resulting from the economics of bloggers—the low entry cost and the highly skewed benefits related to becoming a guru. That is, these two features may induce competing bloggers to resort to “sensational” pieces of information and highly extreme opinions to attract public attention. 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.

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) note, "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 take an extreme negative tone when they want to win the attention war against their competing peers. In this context, competition will inflate extreme and especially negative opinions among competing bloggers. We label the prediction that bloggers will resort to more extreme negative tones in response to competition the *information distortion hypothesis*.

To test these two hypotheses, 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 competitor 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 *tone difference* from its negative mean further in the negative direction by an additional 15%. This negative impact is both statistically significant and economically sizable. Moreover, if we further decompose the analysis into positive and negative tone, we find that competition has no significant impact on positive tone but that competition significantly enhances the magnitude of negative tone; competition also increases the *extremism* of the tone because of its impact on negative tone.

These results are confirmed when we use alternative proxies for competition (e.g., the logarithm of the number of competitors). Additionally, we consider an exogenous event: the change in the number of blog platforms. During our sample period, three new popular blog 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 became more stabilized afterward. This exogenous event effectively represents a sort of structural break in the degree of "potential" competition. Hence, we can directly inspect the impact of competition as new platforms emerge. We find that the increase in the potential competition significantly amplifies the impact of competition for both negative tone and tone difference and renders the tone more extreme. During the peak time, for instance, the impact of the competition dummy more than doubles the average tone difference.

Does the impact of competition on negative tone arise because bloggers are processing more precise negative information or exaggerating 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. While we expect that more information processing is likely to occur among stocks with less public information, guru dreams could induce competing bloggers to exaggerate more for stocks with more public scrutiny—to attract such public attention. Based on three proxies of public scrutiny, including S&P500 affiliation, analyst coverage, and governance quality, we demonstrate that competition among bloggers affects blog tone mostly in firms with high public attention or scrutiny (i.e., high analyst coverage, better governance, and S&P500 affiliation). This pattern suggests that competition exacerbates negative tone rather than encourages more information discovery.

Next, in our second step, we decompose blog tone into the part induced by competition and the part unrelated to competition (i.e., the rest). We find that the part of blog tone that is driven by competition does not have any predictive power in terms of future stock returns. By contrast, the part of blog tone that is unrelated to competition still exhibits significant predictive power for future returns, in terms of both *tone difference* and *negative tone*. These results suggest that competition, far from increasing the informativeness of blogs, raises the negative bias in blogs, which supports the *information distortion hypothesis* as opposed to the *information enhancement 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, 1990, 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). By 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, which illustrates a positive role of social media in terms of information provision, competition appears to distort information and thus weakens the informational contribution of social media. The economics of social media, particularly the part related to information provision, therefore seems to be completely different 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 political polarization that is often observed in public media (e.g., Groseclose and Milo, 2005).

Our work also contributes to the emerging literature on social media. While a vast body of 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 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 cancan 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 important, 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.

We articulate the rest of the paper 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 collect blog information for all the S&P 1500 stocks for 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 the blog posting, and the keywords of the blog. We retrieve from these data all 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, consistent with the gradual 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 3304 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, and the importance of these platforms is reported in the next few columns—in terms of either rank or market shares.[1] 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. Because the two players quickly captured a combined 21% of the market share, some exogenous changes were introduced. Particularly, in these two years, the booming of blog platforms accorded potential bloggers more flexibility in finding a place to express their opinions and thus attracted vast numbers of new participants. Consequently, the degree of competition among all bloggers increased over the same period. Our later tests will use this property to examine the impact of competition.

Our analysis focuses on the following variables. The first set of variables is related to the tone of blogs. We process the linguistic content of each blog by following Loughran and Mcdonald (2008), which allows us to compute the positive and negative tone of a blog article as the 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 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 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 can have non-zero 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 $Blog\_tone\_diff_{i,k,t}$. 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., $\max(Blog\_tone\_pos_{i,k,t}, Blog\_tone\_neg_{i,k,t})$.

[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.

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, tone difference, and degree of extremism for all the blogs covering the same stock in a given month, respectively. We define blog coverage, i.e., $Blog\_coverage_{i,t}$, directly at the firm level as the number of blog articles that are 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 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: $Blog\_tone\_pos_{i,j,t}$, $Blog\_tone\_neg_{i,j,t}$, $Blog\_tone\_diff_{i,j,t}$, and $Blog\_tone\_extreme_{i,j,t}$, which refer the average values for positive tone, negative tone, tone difference, 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 $News\_tone\_pos_{i,t}$, $News\_tone\_neg_{i,t}$, $News\_tone\_diff_{i,t}$, 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 captured directly at the firm level as the number of newspaper articles that are 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 the 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 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.[2]

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 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

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

reported in I/B/E/S and use 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 described in appendix A.

We report 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 quintile 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 bloggers typically write more articles about firms than the top four newspapers write, 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 less negative than newspapers (i.e., newspaper articles have a more negative tone), suggesting that the information that is delivered by blogs is also likely to differ from that provided by newspapers.

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 blog tone difference is positively correlated with DGTW return and that the magnitude of negative blog tone is especially (negatively) correlated with DGTW return. These observations suggest that blogs may contain useful information about stock returns. Of course, whether blogs indeed contain useful information about stock returns needs to be tested in a multivariate specification, which is the task that 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 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 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 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 standard errors at the firm level. The (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 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 to 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 become less involved with the presence of more informed trading in the market. Overall, this table provides preliminary indirect 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 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 tone, 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 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 we cluster the standard errors at the stock and time 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, and to highlight the extent to which blogs can provide information above and beyond public media, we also tabulate the impact of blog tone while controlling for similar newspaper tone. The results indicate that the difference between the positive tone and the negative tone of blogs is highly informative. These results 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). Further, the effect is not only statistically significant but also economically relevant: a one-standard-deviation increase in *Blog_tone_diff* is related to a 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. Hence, both the positive tone and the negative tone of blog articles are generally more informative than public media. By contrast, extremism does not seem to have any predictive power for stock returns. Note that the predictive power of 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 provide. Overall, these results support the *informed guru hypothesis*, indicating that blogs generally tend to be informed rather than to focus on cheap talk.

IV. Competition and Blog Tone

Next, we move on to 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:

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

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 to which the blog tone is extreme (*Blog_tone_extreme*). In addition, $M_{i,j,t}$ 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, in columns (1)-(3), we report the results for tone difference; in columns (4)-(6), we consider positive tone;

[3] 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%.

in columns (7)-(9), we consider negative tone; and in columns (10)-(12), we consider the degree of extremism.

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 *tone difference* further in the negative direction, with the economic magnitude of the impact being approximately 15% of its mean value.[4] Consistent with this negative impact, Models (7) to (9) clearly show 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. By contrast, competition interestingly 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 *information distortion 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 blog tone difference. If we decompose blog tone into positive and negative tone, we see that analyst tone is negatively related to both positive and negative blog tone.[5] This result suggests that the tone of blogs is very different from the tone of professional market watchers, such as analysts. Additionally, the explanatory power of the regression is very high, suggesting that we are indeed identifying the main determinants of blog tone.

We also consider an exogenous event: the change in the number of blog platforms. Three popular blog platforms started in 2007 and 2008. Tumblr was established on Feb. 2007, Movable Type, on Dec. 2007, and Posterous, on May 2008. The emergence of these platforms induced a vast increase in the number of bloggers in 2007 and 2008. To analysis the impact of this exogenous event, we estimate the following specification:

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

where $Peak_t$ is 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 level of the peak dummy variable.

We report the results in Panel A of Table 6 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 tone difference. During the peak time, for instance, the impact of the competition dummy

[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.
[5] 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.

more than doubles the average tone difference.[6] Competition also renders blog tone more extreme. By contrast, in line with our expectations, competition has no impact on positive tone.

V. Blogs and Information

To further confirm the information distortion hypothesis, we must directly investigate whether competition renders blog tone more negative because bloggers provide more precise information or because bloggers simply exaggerate information with a more extreme tone without providing any additional information.

We therefore examine the relationship between blog tone and competition in different sub-samples defined in terms of analyst coverage (*Analyst_num*), governance quality (Aggarwal et al 2009), and SP500 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 governance, and S&P500 affiliation. More specifically, competition exacerbates the negative tone of blogs, especially for stocks that are under high media scrutiny. These results support the information distortion hypothesis.

Finally, we combine our previous results, and we ask whether the link between blog tone and stock returns 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"), and 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, namely, *Blog_tone_diff*, *Blog_tone_pos*, *Blog_tone_neg*, and *Blog_tone_extreme*, and we report the results in Table 8. In columns (1)-(3), we report the results for overall tone; in columns (4)-(6), for positive tone; in columns (7)-(9), for negative tone; and in columns (10)-(12), for the degree of extremism in blog tone.

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 information distortion hypothesis. By contrast,

[6] 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*.

the residual component of blog tone—i.e., the part of blog tone that is not linked to the distortionary effect of competition—predicts future returns, in terms of both tone difference and negative tone. In particular, a one-standard-deviation increase in the residual tone difference (negative tone) predicts a 1.1% (1.09%) annualized abnormal return. [7] The predictive power of the residual tone variables confirms our earlier results that blog tone helps to predict returns.

Conclusion

In this paper, we study the economics of social media based on a unique dataset of blogs. Compared with traditional media, social media is 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 options 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. Thus, the impact of competition on the accuracy of information contained in blogs drastically differs from what we observe in other parts of the economy. For instance, competition improves the accuracy of information supplied by 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.

[7] Similar to Table 3, 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%.

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 that covered the firm in a month
Blog_tone_pos	The average value of the positive tone (weighted value of positive words following Loughran and Mcdonald (2008)) of all articles that covered the firm in a month
Blog_tone_neg	The average value of the negative tone (weighted value of negative words following Loughran and Mcdonald (2008)) of all articles that covered the firm in a month
Blog_tone_diff	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 blog articles that covered the firm in a month
News_tone_pos	The average value of the positive tone (weighted value of positive words following Loughran and Mcdonald (2008)) of all articles that covered the firm in a month
News_tone_neg	The average value of the negative tone (the weighted value of positive words following Loughran and Mcdonald (2008)) of all articles that covered the firm in a month
News_tone_diff	The signed difference between the positive tone and the negative tone of blogs 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 blogs that covered the firm in a month
Other Main Variables	
C2	A variable from Llorente, Michaely, Saar, and Wang (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 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 presents the time series summary statistics for 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, and the importance of the platforms is reported in the next few columns—in terms of either rank or market share. 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 for each year. Our sample covers the period from 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 for the Main Variables

This table presents the summary statistics for our main and control variables. Panel A reports the summary statistics for blog coverage, blog tone, newspaper coverage, and newspaper tone. Panel B reports the summary statistics for blog coverage and tone in the conditional sample, when the firm month has been covered by at least one blog article. In addition, we report the summary statistics for newspaper coverage and tone when the firm month has been covered by at least one newspaper article. Panel C displays the summary statistics for other variables in the following regressions. Panel D reports the Pearson correlation between other firm-month variables in the following regression. All the variable definitions are provided 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

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
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

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 presents the results for the following regression on each stock in a monthly period with firm and month 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 C2 and Flow, for stock i in period $t+1$. C2 is 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 are provided 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 period.

	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 presents the results for the following regression on each stock in a monthly period with firm and 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 performance of stock i in month $t + 1$, (i.e., abnormal return 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.) $Blog\_tone_{i,t}$ 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_diff*), 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. All variables are provided in 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 period.

	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 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:

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

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 to which the blog tone is extreme (*Blog_tone_extreme*). In addition, $M_{i,j,t}$ stacks control variables for stock $i$and fixed effects for blogger j. Panel A uses *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. Panel B uses the 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 are provided in 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 period.

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 the Peak Year Dummy

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: $Blog\_tone_{i,j,t+1} = \beta_0 + \beta_1 \times Competition_{i,j,t} + \beta_2 * Competition_{i,j,t} * Peak_t + C \times M_{i,j,t} + \varepsilon_{i,t+1}$, 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 to which the blog tone is extreme (Blog_tone_extreme). We include Peak_dummy to measure a peak increase in the number of bloggers in 2007 and 2008. We consider an exogenous event: the change in the number of blog platforms. Two popular blog platforms emerged in 2007 and 2008. The emergence of these blog platforms induced an increase in the number of bloggers to a peak in 2007 and 2008. Tumblr was established on Feb. 2007, Movable Type, on Dec. 2007, and Posterous, on May 2008. $\mathrm{Peak_t}$ is a dummy variable that takes a value of 1 in the two years of 2007 and 2008 and 0 otherwise. Panel A uses *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. Panel B uses the 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 are provided in 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 period.

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 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 for each subsample separated by 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_diff*), 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 control variables for stock i and fixed effects for blogger j. Panel A uses *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. Panel B uses the 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 are provided in 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 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_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 Returns

This table presents the results for the following regression on each blogger of each stock in a monthly period with the blogger and month 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"), 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 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 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. In addition, panel B is based 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.

	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