

Blog, Blogger, and the Firm: Can Negative Posts by Employees Lead to Positive Outcomes?

Abstract

Consumer generated media, particularly blogs, can help companies in increasing the visibility of their products without spending millions of dollars in advertising. While a number of companies realize the potential of blogs and encourage their employees to blog; a good chunk of them are skeptical about losing control over to this new media. Companies fear that employees may write negative things about them and this may bring significant reputation loss. Overall, companies show mixed response towards negative posts on employee blogs- some companies show complete aversion, while others show resilience towards negative posts. Such mixed reactions towards negative posts motivated us to probe for any positive aspects of negative posts. In particular, we investigate the relationship between negative posts and readership of an employee blog.

In contrast to the popular perception, our results reveal a potential positive aspect of negative posts. Our analysis suggests that negative posts act as catalyst and can exponentially increase the readership to employee blogs, and suggests that there is some merit to thinking of resilience towards negative posts. Since employees typically write few negative posts and largely write positive posts, the increase in readership to employee blogs generally should be enough to offset the negative effect of few negative posts on employee blogs. Therefore, not restraining few negative posts to increase readership should be a good strategy for the start. This raises a logical question, what should be a firm's policy regarding employee blogging? For exposition, we suggest an analytical framework using our empirical model to answer this question.

Key words: *blog; employee blogs; bloggers; attribution theory; non-linear models; negative posts; influence*

1. Introduction

Companies spend billions of dollars reaching out to their stakeholders – customers, investors, employees and other stakeholders, through advertisements, event sponsorships, and other means. A recent Nielsen Global Survey suggests that blogs are considered a reliable source of information by North Americans and Asians, and are trusted more than all types of advertisements except those in newspapers (Nielsen 2007). Since blogs are trusted more, they can influence customers and other stakeholders in a cost-effective manner (Edelman and Intelliseek 2005). For instance, on March 18, 2006, the Yankee Group/Sunbelt’s survey found Windows to be more reliable than Linux (Eckelberry 2006). However, this finding went unnoticed until Microsoft’s employee Robert Scoble wrote about it on June 6, 2006 (Scoble 2006), after which it generated significant interest¹ - without Microsoft having to pay for the coverage.

Realizing the potential of blogs, many companies encourage their employees to maintain blogs. Jonathan Schwartz, CEO at Sun Microsystems, expresses the importance of employee blogs in these words, *“If you want to lead, blog...We talk about our successes - and our mistakes. That may seem risky. But it’s riskier not to have a blog.”* (Schwartz 2005). IBM and Microsoft, for example, have over 2000 employee blogs (Byron and Broback 2006), and about 10 percent of the workforce at Sun Microsystems maintain blogs (Oliver 2006).

While companies recognize the importance of employee blogs, they also recognize the pitfalls. Employees may not always write positive about their employing companies. Employees may criticize the firm or its products, say embarrassing things about coworkers, or promote competitors. For instance, Michael Hanscom, who used to work at a printshop on Microsoft’s main campus, was fired after he shot some pictures of Mac G5 computers being delivered to the campus and posted them on his blog. Mark Jen, a Google employee, compared Google’s health policies unfavorably to those of Microsoft and suggested that Google provides extensive campus facilities such as free food, car service, dentist, etc. in order to allow employees to work for as many hours a day as possible. He was fired after just 17 days on

¹ After posting the results on March 18, 2006, this finding received no citation for two and half months. But after Robert Scoble’s posting, the survey received more than 37 citations in less than two weeks.

the job. The firing of these employees attracted even more criticism for involved companies (Crawford 2005).

Companies are understandably concerned that negative blog postings by employees may adversely impact the companies' reputation. A typical reaction may therefore be to either completely forbid blogging by employees, or forbid any negative posting whatsoever. But there is an alternative view about negative posts on employee blogs, as explained by Michael Wiley, Director of Communications GM, in an interview with the Wall Street Journal:

“A lot of what blogging is about is authenticity, getting beyond corporate speak and PR, and really creating a conversation. Not being thin skinned and accepting the negatives, that’s key.”

(Brian 2005)

Conventional wisdom and the business press suggest that if readers only see positive posts on employee blogs then they may ignore the blog thinking it to be another marketing outlet deployed by the company. But if an employee candidly discusses flaws in his/her company then readers may consider him to be an honest and helpful person. This perception could lead to increased visibility and credibility of the blog, which could result in a greater exposure to the positive posts contained in the blog as well, the net effect of which may well be an increase in the overall positive influence of the blog on readers (Halley 2003; Scoble and Israel 2006).

In this context, we now explore three questions: (i) Do negative posts help attract more readers to a blog? (ii) If so, why? What might be some of the theoretical underpinnings for this phenomenon? (iii) Finally, how can companies use this knowledge to optimize their employee blogging policies? We next explore each question in more detail, while emphasizing that the main focus and contribution of the paper is with respect to question (i).

For a blog to have any impact, in addition to being perceived as an honest source of information, it must be read by a significant number of readers. Do negative posts attract more readers? To the best of our knowledge, there is no prior work that empirically examines this question. This is a critical question that needs to be answered for firms that are contemplating to let their employees blog. While negative posts may make the blog look more interesting, they may hurt the firm. If negative posts do not attract

more readers, then they are incapable of providing any real benefit to the firm. If negative posts do attract more readers, then there is a chance that the readers will be exposed to the positive posts on the blog as well, which could result in a net overall benefit to the firm. Having a wide readership of employee blogs will thus help the firm use such blogs as a channel for dissemination of credible and timely information.

Using a unique longitudinal sample of employee blogs of a Fortune 500 Information Technology firm, we derive empirical insights into the impact of negative posts on blog readership. We find that with an increase in negative posts, readership increases exponentially up to a certain level of negative posts, and then stabilizes beyond this point. This finding is a key contribution of our paper.

This raises another question – why do negative posts increase readership? Our dataset does not have the requisite granularity to provide a robust answer to this question. However, we draw upon attribution theory from research in social psychology, to present an intuitive explanation of why letting an employee write negative posts may result in increased credibility for that employee’s blog, and why that could translate to increased readership.

We then present an analytical framework that helps identify conditions under which the presence of negative posts could generate greater net positive influence on readers or create more positive views of a firm. For illustration, we apply this analytical framework to data from the Fortune 500 IT firm and calculate the net positive influence of a sample of employee blogs at the company. Our framework helps identify conditions for this company wherein more negative posts by bloggers may result in an increase in the overall positive influence on their readers towards the company.

Our model is of significant interest to firms interested in exploring employee blogging. First, by providing rigorous evidence that negative posts do indeed increase blog readership, our model serves to caution firms that controlling their employees’ blogging behavior may be counter-productive before a point. Second, our model provides a framework for firms to collect and analyze data from their employee blogs, so that the firms can identify what type of blogging activity is most effective in their context, and thereby fine-tune their blogging policies. Our results are of significant interest to IS researchers and social psychologists interested in exploring this emerging phenomenon further.

The rest of the paper is organized as follows. In section 2, we briefly discuss the related literature. In section 3, we lay the groundwork for the empirical analysis, by discussing the data, definitions, and measures. In section 4, we derive an empirical model to test the link between negative posts and blog readership using econometric analysis. In section 5, we use attribution theory to explore why negative posts on an employee's blog might result in increased readership and credibility for the blog. In section 6, we incorporate the empirical relationship derived in section 3 into a framework for companies to evaluate the conditions under which negative posts could increase the overall positive influence of a blog on readers, and we apply this framework to the specific case of the Fortune 500 firm. Finally, in section 7, we conclude with the limitations of this study and future research directions.

2. Literature Review

Our research builds upon and adds to the emerging literature on blogs. Extant literature has studied different facets of blogs. The earliest questions explored by this literature deal with understanding why individuals blog in the absence of any direct incentives (Nardi et al. 2004; Miura and Yamashita 2007; Pedersen and Macafee 2007; Furukawa et al. 2007). These studies find that individuals are driven to document their ideas, provide commentary and opinions, and express deeply felt emotions. Another stream of literature on blogs primarily focus on the volume of posts that are written on a topic in the blogosphere and how does this volume affect their respective variables of interest such as product sales (Goldstone 2006; Onishi and Manchanda 2009; Mishne and Glance 2006), stock trading volume (Fotak 2008), and political event outcome (Adamic and Glance 2005; Farrell and Drezner 2008). They have largely ignored the affiliations of individuals writing blogs, in particular the affiliation by the virtue of being employed by a certain company. Blogger affiliation is an important attribute to consider in blog studies because this can change expectations of readers from a blog radically, and in turn the readership of a blog (Scoble and Israel 2006).

Blogs maintained by employees are of particular interest to employing companies because the business press suggests that employee blogs provide human face to companies, and act as free advertising instruments (Brian 2005; Wright 2005; Halley 2003; Weil 2010; Kirkpatrick 2005). Marketing literature

states that any form of advertising (may it be single advertisement, campaign of advertisements, or a medium of advertising) has two attributes at its core – how many customers you reach, and how strongly you influenced them (Coffin 1963). Companies traditionally focus predominantly on the former core attribute of advertising- audience size (Headen et al. 1977), and in the context of employee blogs this connotes readership. Thus companies that appreciate the potential of employee blogs as an advertising avenue, understandably, want to increase the readership of employee blogs.

There is general paucity of research investigating the readership of employee blogs and the reason for this gap is that the readership data is proprietary, hence not easily accessible. Besides a company's willingness to share data, another challenge is that such data is only available from companies that provide dedicated space on company servers to their employees for blogging. These factors pose a formidable challenge for researchers to study research questions in the employee blogs domain. Our research has only been made possible through a successful data sharing agreement with a Fortune 500 IT firm. Other than our work, very few studies have empirically investigated blogs within an enterprise context (e.g. Yardi et al. 2009, Singh et al. 2010). Yardi et al. (2009) find that when employees blog they expect to receive attention from others. If their expectations go unmet they express frustration and stop blogging. They suggest that firms should help employees receive higher readership for continued blogging. Both these studies emphasize the importance that a firm lays on the readership of its employees' blogs. In comparison to our study, where the employee blogs are accessible to outside individuals the blogs in the above studies are only accessible to the employees. Blogs not accessible outside company completely lose the chance of acting as advertising instruments for a company. It is the blogs that have open access where a firm would be more interested in blog readership and where negative posts by employees may harm it significantly.

Our research is also related to the broader area of user generated content. Specifically, it adds on to the literature that studies how textual characteristics of a post may affect its readership. Recently, Ghose and Iperoitis (2010) found that readers find those reviews to be helpful which are objective, readable, and free of grammatical errors. Lu et al (2010) find that individuals are more likely to trust reviewers whose reviews are moderately objective, comprehensive and readable. Ghose et al (2009) study how the

feedback posted by buyers affect a seller's reputation and his pricing power. In the same vein, our study sheds light on a new aspect, that is, the sentiment of the post and explores its relationship with its readership.

3.Data, Definitions and Measures

3.1 Data

Data for this study come from three archival sources.

(1) The first source is the blog posting and readership data from a Fortune 500 IT firm. This firm is one of the early adopters of Web 2.0 technologies. It provides a platform for its employees to host their blogs. Given that the workforce is primarily technical, blogging as an activity has been widely adopted by its employees. This firm has more than 1000 employees who write blogs on a regular basis. We have access to the time stamped blog posting data along with the post content and the readership data at a blog-week level

(2) The second source is the post citation data from a blog search engine, Technorati.com. Every post has a unique URL (called its “permalink”) that other sites can cite or link to. Search engines such as Technorati provide a count of the number of such citations to each permalink. We collected post citation information for all posts in our dataset.

(3) The third source is the daily XML feed subscription data for each of the bloggers from Bloglines.com. Ask.com also uses the same subscription data set to rank the bloggers according to their popularity (Massie and Kurapati 2006). These data were also collected for all bloggers in our dataset.

3.2 Definitions and Measures

Blog Readership: As discussed earlier the firm would be interested in increasing the readership of its employees' posts. The firm collects the readership data in the form of web server access logs. The firm provided the readership data of a blog to us on a weekly level. We define, $P_{i,t}$, as the number of readers (page-views) who read blog i during week t .

Post Classification: Our key explanatory variable of interest deals with post sentiment. We employed ten graduate students to help us categorize the posts into three categories: positive posts, negative posts, and neutral posts. A post was classified as positive if it promoted the firm (employer)- either the company or any of its products - or if it was critical of a competitor or their products. Conversely, a post was classified as negative if it criticized the firm or its products, or if it promoted competitors or their products. Posts that did neither of the above were categorized as neutral. Every post was categorized by two graduate students, and in case of a tie, the third student's decision was sought. The inter-student reliability for the categorization of posts was 0.91, which suggests a high level of agreement about the category of a post.

Blog Importance: In a blogosphere, a post's importance is measured by the number of other blogs citing it (Leskovec et al. 2007; Adamic and Glance 2005; Drezner and Farrell 2004)². The use of citations as a measure of importance is based on the premise that authors cite what they consider to be important in the development of their work/arguments. A citation captures a considerable amount of latent human judgment, and indicates that the citer is influenced from the citing work and finds it worthy of discussion. Therefore, frequently cited entities are likely to exert a greater influence on readers than those less frequently cited. Therefore our study assumes that when comparing two posts, the one with greater importance is the one being cited more.

Scholars have criticized the assumption that every citation has equal importance (Posner 2000; Pinski and Narin 1976)³. Indeed, it makes little intuitive sense to treat every post as equally important because a citation from Yahoo cannot be put on a par with citation by a site that gets few visitors. Therefore, to

² Citations have been used as a measure of importance in varied settings ranging from scholarly publications (Culnan 1986, 1987; Stigler and Friedland 1975; Medoff 1996; Gittelman and Kogut 2003; Tahai and Meyer 1999; Alexander and Mabry 1994; Ramos-Rodríguez and Ruíz-Navarro 2004), to judicial decisions (Landes et al. 1998; Fred 1992; Kosma 1998), patents (Trajtenberg 1990; Hall et al. 2005; Harhoff et al. 1999; Gittelman and Kogut 2003), and web pages (Brin and Page 1998).

³ Some studies point out, not all citations of an entity may reflect the importance of the entity. Sometimes, entities may be cited because of perfunctory mention and negation (Baumgartner and Pieters 2003). A blogger may cite an article for strategic reasons, as in the case of an article written by his or her senior colleague or supervisor, or may cite a post in order to disagree with it. Nevertheless, when a blogger cites another blog post to disagree with it, this is also a gauge of the importance of the post, since the blogger could have ignored the post as unimportant instead of explaining why it is inaccurate or why he takes an opposite stance. In any event, critical citations do not pose a problem in our data set, where they amount to less than 5.1% of citations. Although these limitations are important, using citations as a measure of influence is less prone to systematic biases than other measures.

account for this difference, we assign weights to each citation based on the number of citations to the citing site. Therefore, in order to measure the weighted citations to a post ($W_{p,t}$), we consider the citations, and weigh them with their 2nd level citations (Please see an Appendix A to this paper for an example calculation). This weighted citation of a post represents its importance in our study.

So far we have defined the importance of a post and how to measure it. A blog can have more than one type of post: positive posts will cast the company in a positive light, and negative posts will leave the reader with a negative impression of the company. We define $W_{i,t}^+$, $W_{i,t}^-$, and $W_{i,t}^0$ as weighted citations received by positive, negative and neutral posts respectively. Each of these measures is calculated by summing the weighted citations of the posts of its type displayed on the blog during period t . We also construct the measure weighted citations to a blog as the sum of weighted citations of posts on a blog ($W_{i,t} = W_{i,t}^+ + W_{i,t}^- + W_{i,t}^0$), without regard to whether the posts are positive, negative, or neutral. The weighted citations capture two effects- traffic driven to a blog purely due to higher visibility brought by links; traffic driven to a blog due to the quality and the importance of the blog in the blogosphere.

Ratio of Negativity: In order to measure the extent of negative posts on each blog, we define the ratio of negativity as the ratio of negative content to positive content on a blog and measure it as ($R_{i,t} = W_{i,t}^- / W_{i,t}^+$). The ratio of negativity indicates the extent to which the negative posts on the blog are considered important in the blogosphere compared to positive posts on the same blog. As ratio of negativity increases, it implies that blogger is posting more influential negative posts or equally influential but more negative posts. The effect of ratio of negativity on readership is of key interest to us. A positive relationship between ratio of negativity and blog readership would indicate that negative posts lead to higher readership.

Blog Popularity: The readership of a blog can be affected by the popularity of the blog. To control for a blog's popularity, we use two blog-level variables to measure the popularity of the blog during past periods: page-view lag (Battelle 2005) and XML feed subscription data (Massie and Kurapati 2006). The rationale for using page-views lag is that a more popular blog should have more page-views during past periods. Page-views lag can also serve as a proxy for the blog quality and employee-specific latent variables that make some blogs more popular than others (Wooldridge 2001). "XML feed subscribers"

refers to the number of people who subscribe to a blogger's posts. Whenever this blogger publishes a post, the headline of the post is sent to the subscriber automatically. The number of people subscribing to a blog is likely to be proportional to the popularity of a blog, and so we use XML feed subscribers as a measure of popularity of the blog (Massie and Kurapati 2006). The variable definitions are provided in Table 1.

-----[Insert Table 1 Here]-----

We randomly selected 211 employees, and collected data for 13 weeks starting from January 2007, which resulted in a total of 2,743 employee-week observations. We split the data set into two subsets: the first 11 weeks' data (2,321 observations) were used as the training dataset, to estimate the model parameters; the last 2 week's data (422 observations) were used as a test dataset to illustrate the fit and compare our model with alternate specifications. Table 2 provides the descriptive statistics for key variables. Our unit of observation for statistical analysis in this study is a blog-week. In a week, on average, 2402.77 visitors visited a blog in our sample. On average a blogger posted once every two and a half days. A post was archived when fifteen subsequent posts had been made on the blog. Bloggers on average posted less than one negative post in every set of fifteen posts. Therefore, once a negative post was made, it stayed on the blog page for about forty days. As a result, there were many days during the testing period when the blogs displayed negative posts. From the descriptive statistics, one can make the following insightful observation:

- The total number of weighted citations received by positive posts is higher than the weighted citations received by negative posts. The number of weighted citations received by the positive posts on average is (59.36) and that by negative posts is (12.31). Note that weighted citations of a type of post is calculated as the sum of the weighted citations of all posts of that type that are displayed during the focal week.
- Negative posts receive disproportionate citations from others. We found that the extent of weighted citations received by a negative post is, on average, approx double the extent of weighted citations received by a positive post (see Table 2). Note that the weighted citation is calculated at week level. In a given week on average a blog displays 0.97 negative and 7.24

positive posts. Hence, per post the numbers of weighted citations for negative and positive posts are $(12.31/0.97=12.69)$ and $(56.56/7.24=7.81)$.

-----[Insert Table 2 Here]-----

4. Econometric Specifications

4.1 Blog Readership Model

We now address the task of empirically examining whether negative posts increase blog readership. We start with a simple linear specification where the page views ($P_{i,t}$) is affected by blog popularity ($P_{i,t-1}$, and $S_{i,t}$), blog importance ($W_{i,t}$), ratio of negativity ($R_{i,t}$) and the interaction of blog importance with the ratio of negativity. Consider the following linear model specification:

$$P_{i,t} = \alpha_0 + \alpha_1 P_{i,t-1} + \alpha_2 W_{i,t} + \alpha_3 W_{i,t} R_{i,t} + \alpha_4 R_{i,t} + \alpha_5 S_{i,t} + \varepsilon_{i,t} \quad (1)$$

If this specification is appropriate, pooled OLS will give consistent and efficient estimates. This specification includes an interaction term which has the potential to cause multicollinearity problems. We tested for the presence of multicollinearity in our data. The maximum Variance Inflation Factor (VIF) in specification (1) is 1.9 , which is much less than 10 indicating that multicollinearity is not a problem. Furthermore, if a specification has an interaction term and there is a multicollinearity problem between the independent variables and the interaction term, then typically the interaction term comes out to be non-significant (Jaccard and Turrisi 2003). However, in this case, it comes out to be strongly significant; hence this serves as another check for the conclusion that multicollinearity is not a problem here. Table 3 presents the estimated results (pooled OLS) from the training data set and reports that other than XML feed subscription and Ratio of negativity, all other variables significantly influence a blog's readership.

The above analysis tells us that

- The Ratio of negativity does not directly affect the blog readership but moderates the effect of importance of a blog on readership.
- The XML Feed Subscription is non-significant. After controlling for the popularity of a blog through page-view lag, XML Feed Subscription does not significantly increase the explanatory power of the model.

-----[Insert Table 3 Here]-----

Another possible specification (2) is to decompose the weighted citation of a blog on a reader into the citations to the positive, negative, and neutral posts of a blog, and their interaction with the ratio of negativity.

$$P_{i,t} = \alpha_1 P_{i,t-1} + \alpha_2 W_{i,t}^+ + \alpha_3 W_{i,t}^- + \alpha_4 W_{i,t}^0 + \alpha_5 W_{i,t}^- R_{i,t} + \alpha_6 W_{i,t}^0 R_{i,t} + \alpha_7 R_{i,t} + \alpha_8 S_{i,t} + \varepsilon_{i,t} \quad (2)$$

Results for specification 2 (estimated as pooled OLS) are presented in Table 4. The results are consistent with those from specification 1. Additionally, these results tell us that:

- For all three types of posts, the coefficient corresponding to weighted citations is positive and significant. This confirms our belief that the more important posts receive higher readership.
- Further, both the interaction term coefficients are positive and significant indicating that important posts receive more readership when displayed on a blog with negative posts than otherwise.

-----[Insert Table 4 Here]-----

While specifications (1) and (2) are very simple and hence preferable, additional analysis reveals that they fail on several econometric issues. Neither of the two specifications capture non linearity in the data. Using Ramsey's RESET test, we reject the null hypothesis ($p < 0.009$) that all non-linearity patterns are captured in the specification. Note that the interaction term in specification (1) is linear. To illustrate that there is a need for non linear interaction term we present Figure 1.

-----[Insert Figure 1 Here]-----

Figure 1 is a plot between the rate at which readership increases per citation and the Ratio of Negativity. It shows that there is a non-linear pattern (initial exponential increase followed by stability) instead of a straight line. This non-linearity is the reason why Ramsey's RESET test failed. We find that such curves are best modeled using the Michaelis-Menten model of the kinetics of chemical reactions (Seber and Wild 2003; Bates and Watts 1988; Michaelis and Menten 1913). The influence of negative posts on the readership for a blog appears to share some similarities with the chemical reaction in a

presence of catalyst⁴, for which a nonlinear model such as the Michaelis-Menten model is appropriate. Negative posts appear to be most effective when present in small quantities: readership increases with the ratio of negativity exponentially at first, then at a rapidly declining rate, and finally stabilizing after a point.

Based on this observation, we changed our model specification to the following.

$$P_{i,t} = \beta_1 P_{i,t-1} + \beta_2 W_{i,t} + \beta_3 W_{i,t} \frac{R_{i,t}}{R_{i,t} + \beta_4} + \beta_5 R_{i,t} + \beta_6 S_{i,t} + \varepsilon_{i,t} \quad (3)$$

Using Ramsey's RESET test, this time we failed to reject the null hypothesis ($p > 0.37$) which indicates that all the non-linearity in the data is captured. Results are given in Table 5 (estimated as pooled OLS). Even after accounting for the non linearity in the data, XML feed subscriptions does not significantly affect readership. While, the ratio of negativity does not have a significant direct effect on the readership, it has a non linear significant effect on readership, whereby readership increases with the ratio of negativity exponentially at first, then at a rapidly declining rate, and finally stabilizes after a point.

-----[Insert Table 5 Here]-----

While the results from specification (3) clearly indicate that negative posts may not always have a negative impact on blog readership, several other econometric issues need to be addressed to gain more confidence in our results. To account for other econometric issues, we modify specification (3) in several ways. First we incorporate blog specific unobserved effects in the specification. Second, we allow first order serial correlation among errors in the specification. We test for higher order serial correlation also and find that the first order serial correlation is appropriate.

$$P_{i,t} = \beta_1 P_{i,t-1} + \beta_2 W_{i,t}^+ + \beta_3 W_{i,t}^- + \beta_4 W_{i,t}^0 + (\beta_5 W_{i,t}^+ + \beta_6 W_{i,t}^- + \beta_7 W_{i,t}^0) \frac{R_{i,t}}{R_{i,t} + \beta_8} + \beta_9 R_{i,t} + \beta_{10} S_{i,t} + \beta_{11i} + \gamma_{i,t}$$

$$\gamma_{i,t} = \eta \gamma_{i,t-1} + \epsilon_{i,t}, \quad \epsilon_{i,t} \sim N(0, \sigma^2) \quad (4)$$

⁴ Negative posts appear to play a role similar to that of a substrate in the presence of catalyst in a chemical reaction: page-views (reaction rate) increases with increasing proportion of negative posts (substrate concentration), asymptotically.

Finally, we note that there could be a possible simultaneity in readership and importance of a blog for the same period: if readership for a post increase during a period, then that post is likely to get more citations, which increases the importance of a blog. We require an instrument variable that is correlated with the importance of a blog but not with sudden changes ('shock') in readership that would reflect idiosyncratic errors. A lag in the weighted citation of a blog can be a valid instrument if the model is dynamically complete or sequentially exogenous (Wooldridge 2001), i.e., if the shock in readership during a given period does not change the weighted citations of a blog in past periods. The econometrics literature has shown that if the dynamic completeness condition is violated, there will be autocorrelation in idiosyncratic errors. We tested for autocorrelation after accounting for first order serial correlation (Wooldridge 2001) but failed to reject the null hypothesis ($p > 0.71$), suggesting that the assumption of dynamic completeness in our model is justified. Therefore, we can use lags of weighted citations of a blog as valid instruments for importance of a blog. We choose a two period lag and a two period lagged first difference of weighted citations of a blog as two instruments. The two instruments provide us the latitude to perform additional weak instrument tests. On surface, both the instruments appear to satisfy both the conditions for being a good instrument. First, two period lag and lagged first difference is correlated with the present weighted citations of a blog because both periods share many posts in common. Secondly, the lags may not be correlated with the shock in readership because the shock in readership occurs in the present period.

We performed a two-stage regression estimation. In the first stage, the importance of a blog is estimated as a function of a two period lagged weighted citations, the lagged first difference weighted citation and other exogenous variables. Note that this corresponds to three first stage equations, one each for positive, negative and neutral importance. Predicted values of the importance of a blog are calculated from the estimates of this regression. In the second stage, the readership is estimated as a function of these predicted importance and other exogenous variables as specified in specification (4). The standard deviation of the parameters is adjusted appropriately. The second stage is estimated as a maximum likelihood estimation (Davidson and MacKinnon 1993). The blog specific unobserved term is estimated non-parametrically to avoid the bias which may result if one specifies a wrong parametric functional form

for them (Heckman and Singer 1984). As a first indicator of instrument strength, the F-statistic of the three first stage regressions are 47.59 for positive posts, 37.15 for negative posts, and 29.31 for neutral posts, which indicate that the instruments are not weak (Davidson and MacKinnon 1993). The assumption that instruments are not correlated with the error terms cannot be directly tested, but can be tested indirectly if the model is over-identified, that is the number of instruments is more than the number of endogenous variables (Stock et al. 2002). Since our model is over-identified, we perform two tests to check if the instruments are truly exogenous. First, we perform the Sargan test (Arellano and Bond 1991; Stock et al. 2002). The Sargan test involves regressing the second stage IV residuals \hat{u}_i^{IV} on all instruments and exogenous variables. An R^2 from this regression is obtained. The test statistic is then $S=nR^2$ where n is the number of observations. Under the null hypothesis that all instruments are exogenous S is distributed as χ_1^2 . We fail to reject the null that the instruments are exogenous: (positive) $p=0.37$; (negative) $p=0.55$; (neutral) $p=0.31$. The second test that we employ is a J-test (Stock et al. 2002). In the J-test, IV residuals are regressed on the instruments and other control variables; an F-statistic is calculated from this regression and the test statistic (F) is distributed as χ_1^2 under the null of exogenous instruments (Stock et al. 2002). We again fail to reject the null hypothesis of all instruments being exogenous at: (positive) $p=0.36$; (negative) $p=0.56$; (neutral) $p=0.31$.

While we have assumed the blog specific unobserved effects to be uncorrelated with other explanatory variables. Under this assumption it is appropriate to treat these effects as random effects. However, if these blog specific unobserved effects are correlated with other explanatory variables our results would be biased. To address this issue, we estimate another regression where we treat the blog specific unobserved effects as fixed effects. The second stage regression is transformed by rho differencing followed by first differencing to get rid of serial correlation in errors and the unobserved fixed effects. The resulting dataset is estimated through a maximum likelihood estimation. The estimates obtained from this fixed effects regression are tested against the estimates from the random effects estimation by a Hausman test (Hausman 1978). We fail to reject the null that both the estimates are consistent at $p=0.49$. Hence, the results with random effects are consistent as well as efficient. Further, we find that the inclusion of time dummies while improves the log likelihood of the model, the

specification without time dummies has a better Bayesian information Criterion (BIC) and Akaike Information Criterion (AIC). (Davidson and MacKinnon 1993; Wooldridge 2001). Table 6 presents the three sets of results (1) specification (4) with blog random effects and AR1 error structure (2) specification (4) with blog random effects, AR1 error structure, and time dummies, (3) specification (4) with blog fixed effects and AR1 error structure.

-----[Insert Table 6 Here]-----

The results from Table 6 clearly highlight the way in which ratio of negativity affects blog readership. It shows that the readership of a blog would increase exponentially initially with the ratio of negativity, then at a declining rate, after which it stabilizes. From these results it is obvious that a moderate amount of negative posts would be beneficial for a blog.

4.2 Model Comparison

We consider the following three model forms and contrast the results.

- An initial exponential increase followed by stability

This is the case when, *ceteris paribus*, the increase in the ratio of negativity initially leads to an exponential increase in readership, after which readership does not change. We take two model specifications for this case: a base model that uses the Michaelis-Menten formulation and a model that uses the exponential function (Seber and Wild 2003; Bates and Watts 1988).

- An initial exponential increase followed by a decrease

This is the case when, *ceteris paribus*, the increase in the ratio of negativity on a reader leads initially to an exponential increase in readership, after which readership starts decreasing as the ratio of negativity increases. We consider two model specifications for this case: a refinement of the base model that uses the Michaelis-Menten formulation, and the quadratic growth model (Seber and Wild 2003; Bates and Watts 1988).

- Linear increase

This will be the case when, *ceteris paribus*, the increase in the ratio of negativity leads to a linear increase in readership.

We also tested for logarithmic specification but it performed the worst of all the approaches considered. For brevity's sake, we do not report the model based on logarithmic specification. The model selection is based on two widely accepted model selection criterion: BIC and AIC. (Davidson and MacKinnon 1993; Wooldridge 2001). Models are shown in Table 7 and the empirical comparisons are shown in Table 8. Results indicate that our base model, specification (4), based on the Michaelis-Menten model of chemical kinetics offers the best fit.

-----[Insert Table 7 and 8 Here]-----

4.3 Blog Citation Model

In the above analysis, we have shown that the blogs with negative posts have higher readership compared to blogs with positive posts only. We now focus our attention on whether negative posts also get higher citations as compared to positive posts. Our citation specification is given in (5)

$$C_{i,p,t} = \beta_1 P_{i,t-1} + \beta_2 Post_{i,p,t}^+ + \beta_3 Post_{i,p,t}^- + \beta_4 R_{i,t} + \beta_5 S_{i,t} + \beta_{6i} + \gamma_{i,t} \quad (5)$$

In this specification, C_{it} is the number of citations received by a post p (posted at time t) by blogger i within 3 weeks of the posting time t . $P_{i,t-1}$ is as before the lagged readership of the blog, $Post_{i,p,t}^+$ is an indicator variable which equals 1 if the post is positive and zero otherwise, $Post_{i,p,t}^-$ is an indicator variable which equals 1 if the post is negative and zero otherwise. Specification (5) is at post level. Since $C_{i,p,t}$ is a count variable, specification (5) is estimated as a Negative binomial random effects model. Maximum likelihood estimation is used and blog specific unobserved random effects are accounted through a non-parametric distribution. Results for specification (5) are provided in Table 9.

-----[Insert Table 9 Here]-----

Results reveal several interesting insights. Negative posts receive significantly higher number of citations compared to positive and neutral posts. Difference between the number of citations received between the positive and neutral posts is insignificant. Posts on blogs with higher ratio of negativity receive more citations. These results indicate that negative posts not only increase the readership of a blog but they also increase the citations of its other posts. To test this result more rigorously, we estimate the specification (6) which allows interactions of post type with ratio of negativity:

$$C_{i,p,t} = \beta_1 P_{i,t-1} + \beta_2 Post_{i,p,t}^+ + \beta_3 Post_{i,p,t}^- + \beta_4 R_{i,t} + \beta_5 Post_{i,p,t}^+ R_{i,t} + \beta_6 Post_{i,p,t}^- R_{i,t} + \beta_7 S_{i,t} + \beta_8 i + \gamma_{i,t} \quad (6)$$

Results for specification (6) are shown in Table 9. It reveals that the interaction terms are significant. This confirms the earlier finding from specification (5) that negative posts increase citations for future posts. Interestingly, the coefficient for interaction between positive post dummy and ratio of negativity is significantly higher than the interaction between negative post dummy and ratio of negativity ($p < 0.01$). However, on average the citations received by the positive post are still lower than the citations received by a negative post. Hence, a negative post on the blog help a positive post receive much more citations compared to other type of posts. Positive posts typically receive fewer citations than the negative posts but their citations increase significantly if they are posted by a blogger who posts negative posts also. We also tested the specifications (5) and (6) by replacing citations by weighted citations and the results are consistent.

5. Theoretical Discussion

Whereas the discussion in Sections 3 and 4 focused on establishing the relationship between negative posts and readership, the present section focuses on the possible reasons for this relationship. Visitors to a blog can be classified as new visitors and loyal readers. There are certain characteristics of a blog that may affect the likelihood that an individual who has never visited the blog would visit it. A highly visible blog would receive a greater amount of new visitors. Negative posts affect the visibility of a blog. As shown earlier, negative posts receive higher number of citations from others increasing their visibility in the blogosphere. Negative posts also increase the citations received by other types of posts which increases the visibility of the overall blog in the blogosphere. This increased visibility of a blog helps in attracting new readers. Note that the traffic that is driven to the blog through visibility is appropriately captured through the blog importance in Specification (4). This also highlights that negative posts have an effect on readership above and beyond the visibility effect as the even after controlling for the weighted citations, Ratio of negativity has a significant effect on blog readership. Therefore, another major effect of negative posts on readership is through converting new visitors to loyal visitors.

To explain how negative posts help convert some of the new visitor to loyal visitors we view the phenomenon through the lens of attribution theory. Before we proceed further, we want to emphasize that the limitations imposed by our data do not permit us to empirically test the applicability of attribution theory, which we leave for future research.

In a social influence situation, attribution theory helps to explain how people attribute a cause to somebody's behavior and what the results of those inferences are (Jones et al. 1987). An antecedent that people use in making attributions is the perceiver's belief about what others would do in the same situation (Kelley and Michela 1980). Many studies on blogs suggest that readers expect employee bloggers to promote their firms (Scoble and Israel 2006; Edelman and Intelliseek 2005; Byron and Broback 2006; Halley 2003; Brian 2005). But if employees blog about the failures of their employer or promote positive aspects of competitors this may contradict a reader's preconceived belief. Kelly's augmentation principle states that a perceiver observing an action contradicting her belief will attribute that action more to the actor's disposition than to situational pressures (Kelley 1972). Consider the following example: Yoda is a GM employee who blogs at gmblog.com and declares his affiliation on his blog. He chooses to declare the problems in their hybrid models, along with pointing out how GM has the best cars in other segments (positive posts) and writing about his trip to Switzerland (neutral posts) on his blog. A negative post contradicts a reader's, say Cindy's, prior belief about an employee blog. She expects a company spokesperson to write in favor of his company (Folkes 1988), but the disconfirmation of this expectancy leads her to attribute the blog primarily to factual evidence. The more she attributes the blog to factual evidence rather than to Yoda's affiliation, the more she attributes Yoda's action to his honesty (Wood and Eagly 1981). While, some amount of negative posts can increase credibility, a huge amount of negative posts on a blog may have an adverse effect. Readers may discard the blogger considering him to be a disgruntled employee.

We now consider the other side of the causal coin, namely the consequences of the attributions made by a reader about the blogger on the reader's subsequent actions. Consumer research studies explain that when a perceiver, who ascribes a helpful act to the actor's disposition rather than to environmental factors, feels gratitude towards the actor and this in turn instills "person loyalty" (Weiner 2000).

Continuing with our example, from the above discussion, Cindy is more likely to make attributions of honesty and helping nature about the personal disposition of Yoda in the light of his choice and perceive that Yoda helped her by providing full information even at the risk of upsetting his employer; therefore, she is likely to convert to a loyal reader as well as to recommend others Yoda's blog increasing his readership.

6. Analysis of Firm Policies for Employee Blogging

Any form of advertising (may it be single advertisement, campaign of advertisements, or a medium of advertising) has two attributes at its core – how many customers you reach (in this context readership), and how strongly you influenced them (Coffin 1963). Next we provide a framework focusing on the second core attribute of advertising- how negative posts influence readers.

In this section, we model how the functional form suggested by the Michaelis-Menten specification can help a firm to decide when employees should be allowed to write negative posts. As discussed earlier, one way that employee blogs may affect a firm is through influencing a reader's perception about the firm. A firm typically faces a tradeoff where allowing negative posts may influence the reader's perception negatively, but negative posts also increase blog readership overall and the citations received by positive posts. Faced with such a problem, a firm can completely forbid its employees to post anything negative, or leave it to their discretion but impose an upper bar on the level of negativity allowed on a blog. If a firm allows employees to post anything, the net positive influence of a blog i at time t on readers is given by:

$$\text{Net Positive Influence on readers} = NPI_n = (I_{i,t,n}^+ - I_{i,t,n}^-) * P_{i,t,n} \quad (7)$$

The subscript n indicates that negative posts are allowed. In this equation, $P_{i,t,n}$ is given by specification 4. $I_{i,t,n}^+$ ($I_{i,t,n}^-$) is the influence of positive (negative) posts. Based on the literature on social influence, we define influence of a post as measured by a function of its weighted citations⁵ (Burt 1998). The influence I of a post can be expressed as $I=W^\nu$ where ν is the power law coefficient, and W represents the weighted citations for the post. A value of ν close to one indicates that the post is only as influential as

⁵ Firm can also measure the influence of post using sales data.

its weighted citation. Whereas, a value much less than one implies that each post is equally influential irrespective of its weighted citations, a value much larger than one implies that as number of weighted citations increases, the influence of a post increases exponentially. As one may observe this relationship is quite flexible. The power law coefficient can be allowed to vary for different types of posts (positive, negative, neutral). For example, in the analysis, we have allowed it to differ for negative (vn) and positive (vm) posts.

In the Michaelis-Menten based formulation, each variable applies to an individual blogger on a weekly basis. However, for simplicity we now replace each variable in specification (4) and (7) with its average value across bloggers and across time to yield the average net positive influence on readers for a firm not imposing any restrictions on posting. This averaging of variables is justified considering that bloggers allowed to post negative entries will periodically overshoot and undershoot the optimal level of negative posts. As long as the average level of negative posts across bloggers and across several periods is close to the optimal level, the purpose of the decision is served. Therefore, we have Net Positive Influence on readers when negative posts are allowed,

$$\overline{NPI}_n = (\overline{I}_{i,t,n}^+ - \overline{I}_{i,t,n}^-) * \overline{P}_n \quad (8)$$

where

$$\overline{P}_n = \frac{1}{1-\beta_1} \left(\beta_9 + \beta_2 \overline{W}_n^+ + \beta_3 \overline{W}_n^- + \beta_4 \overline{W}_n^0 + (\beta_5 \overline{W}_n^+ + \beta_6 \overline{W}_n^- + \beta_7 \overline{W}_n^0) \left(\frac{\overline{W}_n^-}{\overline{W}_n^- + \beta_8 \overline{W}_n^+} \right) \right) \quad (9)$$

Simplifying from equations (8) and (9) gives the average net positive influence on readers for a firm allowing negative posts in terms of publicly available data.

$$\overline{NPI}_n = \frac{(\overline{W}_n^{+vm} - \overline{W}_n^{-vn})}{1-\beta_1} \left(\beta_9 + \beta_2 \overline{W}_n^+ + \beta_3 \overline{W}_n^- + \beta_4 \overline{W}_n^0 + (\beta_5 \overline{W}_n^+ + \beta_6 \overline{W}_n^- + \beta_7 \overline{W}_n^0) \left(\frac{\overline{W}_n^-}{\overline{W}_n^- + \beta_8 \overline{W}_n^+} \right) \right) \quad (10)$$

Note that in this equation, we have allowed the power law coefficient to differ for positive (vm) and negative (vn) posts. If a firm doesn't allow its blogger to post anything negative, then the average net positive influence on readers can be calculated as:

$$\overline{NPI}_p = \frac{(\overline{W}_p^{+vm})}{1-\beta_1} (\beta_9 + \beta_2 \overline{W}_p^+ + \beta_4 \overline{W}_p^0) \quad (11)$$

Subscript p indicates that negative posts are not allowed.

Firm can compare the two options and choose to allow negative posts when $\overline{NPI}_n > \overline{NPI}_p$

Proposition: There exists a cutoff value of the average negative influence \bar{I}^{-*} such that (a) for all values of the average negative influence on a reader $\bar{I}^- \in (0, \bar{I}^{-*})$, the decision of allowing negative posts creates more average net positive influence on readers than the decision of no negative posting, whereas (b) for $\bar{I}^- > \bar{I}^{-*}$, the decision of no negative posting offers a higher average net positive influence on readers than the decision of allowing negative posts.

(Please see Appendix B for a Proof of this Proposition).

This proposition suggests that prohibiting negative posts is not always the best strategy. The basic intuition is that when a blog has a small number of negative posts, it attracts more readers, who are also exposed to the (more numerous) positive posts on the blog. As more negative posts are added, more readers are attracted at a decreasing rate (as shown empirically in the earlier section), and each of these readers is exposed to a smaller proportion of positive posts. At some point, the benefit of having negative posts in order to increase net positive influence on readers starts to disappear. Eventually, a sufficiently large number of negative posts leads to a worse net positive influence on readers than having no negative posts at all, making the decision of no negative posting better. Please note that a sufficient condition for the above proposition to hold is provided in an appendix to this paper. Figure 2 shows a comparison of the net positive influence on readers generated by the decision of allowing negative posts and prohibiting negative posts.

-----[Insert Figure 2 Here]-----

Diverse Reader Types. It is possible that readers who visit the blog through a citation only read that post and no other post. In such scenarios the reader would form that impression of the firm which is displayed by that one post. To account for such behavior, let us assume a fraction $f_1(f_2)$ of readers coming through a citation and reading only the cited positive (negative) post. For simplification, further assume that all other readers read all posts that are displayed on the blog during their visit. Then the fractions of readers who read all displayed posts on the blog are $(1-f_1-f_2)$. Under these assumptions, the net positive influence from allowing negative posts is:

$$\overline{NPI}_n = [\alpha_p \overline{W}_n^{+vm} f_1 - \alpha_n \overline{W}_n^{-vn} f_2 + (\overline{W}_n^{+vm} - \overline{W}_n^{-vn})(1 - f_1 - f_2)] * \overline{P}_n, \quad (12)$$

where $0 \leq \alpha_p \leq 1$ and $0 \leq \alpha_n \leq 1$ are constants that account for the fact that only one post is read by the reader and adjust the influence appropriately. Note that there are 15 posts displayed on a blog at any time and \overline{W}_n corresponds to the sum of the weighted citations of those 15 posts. For example, out of these posts on average in our data only 7.24 posts are positive leading to $\alpha_p = 1/7.24$. α accounts for the fact that the reader is reading only one of those posts and hence allows influence from only that post. Similarly the net positive influence from forbidding negative posts is:

$$\overline{NPI}_p = [\alpha_p \overline{W}_p^{+vm} f_1 + (\overline{W}_p^{+vm})(1 - f_1)] * \overline{P}_p. \quad (13)$$

A firm can compare both these strategies and choose to allow negative posts if $\overline{NPI}_n > \overline{NPI}_p$. Certain interesting insights can be revealed by comparing it to equations (10) and (11). First, as the fraction f_2 increases, the cutoff point under which allowing negative posts is optimal (I^*) shifts towards zero. A simple explanation of this phenomenon is that as f_2 increases, a large fraction of readers who are visiting the blog are reading only negative post and hence are exposed to negative influence. And there would be smaller fraction of readers coming through links to negative posts and getting exposed to positive posts displayed on the blog also. Second, for $\alpha_p > 1 - \overline{W}_n^{-vn} / \overline{W}_n^{+vm}$, an increase in fraction f_1 shifts the cutoff point, under which allowing negative posts is optimal, away from zero. Below this limit, an increase in fraction f_1 shifts I^* towards zero. To better understand it let us restate the condition as $\overline{W}_n^{-vn} > (1 - \alpha_p) \overline{W}_n^{+vm}$. The left hand side of this condition represents the influence from negative posts. The right hand side represents the influence from positive posts that is not exposed to readers who read only the cited positive post (belong to fraction f_1). When α_p satisfies this condition, it implies that there are very few positive posts on the blog. The additional influence that the reader could be exposed to by reading all positive posts is less than the influence from negative posts.

6.1 Application of the Framework to the Case of Our Research Setting

Firms that encourage blogging by their employees can easily access (and are accessing) the kind of data required for the above analysis. We now apply our analysis to data from the Fortune 500 company and illustrate how it can provide suitable recommendations for optimal blogging policy. Using specifications

(10) and (11), we calculate the upper limit of average negative influence, $\bar{I}^{-*} = 46$ ($R=0.81$) (for $vn=vm=1$). Beyond this point, employee blogs may still generate net positive influence on readers under the decision of allowing negative posts, but can generate more net positive influence on readers if employees refrain from any further negative posts about the firm. In our data set we find that the actual average negative influence on a reader is $\bar{I}^{-} = 12.31$ ($R=0.22$) (for $vn=vm=1$), which is less than the calculated upper value. Furthermore, the optimal average negative influence on a reader for the whole data set, $\bar{I}^{-**} = 15.1$ is ($R=0.27$) (for $vn=vm=1$), which is more than the present average negative influence on a reader. Figure 3a plots how \bar{I}^{-*} and \bar{I}^{-**} change as vn and vm vary. As we discussed earlier, at $vm(vn)$ fixed as $vn(vm)$ increases both \bar{I}^{-*} and \bar{I}^{-**} decrease (increase). Similarly, using Specifications (12) and (13), we plot \bar{I}^{-*} and \bar{I}^{-**} for varying f_1 and f_2 (constant $vn=vm=1$). Results are presented in figure 3b. For a given f_1 as f_2 increases \bar{I}^{-*} and \bar{I}^{-**} decrease. It is important to note that \bar{I}^{-*} and \bar{I}^{-**} vary widely as vn , vm , f_1 or f_2 vary. Hence, a firm should pay attention in estimating these parameters. A firm can estimate these parameters by combining the blog data with the sales data.

-----[Insert Figures 3a and 3b Here]-----

Therefore, the sample data set and the results based on our suggested framework suggest that the concerned firm's decision of allowing its employees to write negative posts is good only if the true values of vn , vm , f_1 and f_2 lie in certain ranges as illustrated by Figures 3a and 3b. In summary, in order to determine whether to relax or tighten its blogging policy, the firm needs to know how far removed its employee blogs may be from the optimum.

We realize that while a firm can decide on the extent of negative posts to be allowed, exactly how a firm implements its desired blogging policy is beyond the scope of this paper. We are not advocating, for instance, that a firm should actively encourage its employees to post negative comments on their blogs. Firms may not choose to restrict certain types of negative posts, which is different from actually encouraging negative posts. Examples of actual blogging policies of firms suggest that some firms set broad guidelines that do not forbid critical posts by employees, as long as those posts are written carefully – posts that are not rude or insulting, that don't reveal company secrets, and don't violate the law.

7. Conclusions and Future Research Directions

One of our main empirical findings is that, *ceteris paribus*, negative posts exponentially increase the readership initially and after a point the page-views do not increase. We also present some theoretical grounding for the idea that negative posts do not always harm a firm and under some conditions could create more net positive influence for the firm. Firms should therefore be careful when discouraging or prohibiting negative posts by their employees. Firms can decide whether to allow or prohibit negative posts using our suggested framework. However, the exact process by which a company should regulate its employee blogs is beyond the scope of this paper. Here we have considered only the case of a uniform decision of allowing or prohibiting negative posts for all employees; nevertheless, our suggested framework is fairly general and the decision can be customized for bloggers in specific firms. For example, bloggers may be heterogeneous in their influence owing to several factors such as job title and experience. This can be allowed by having a blogger-specific influence coefficient.

We do not want to overstate our results. Our analysis has a few limitations. A limitation of our analysis is that net positive influence on readers, i.e., overall positive effect of a blog on readers towards the blogger's employing firm may not directly map to monetary profits or losses for the firm. It would be difficult to map the effect of employee posts on blogs to product sales because such a study would need to control for all the other plausible causes, such as the firm's advertising, competitors' advertising, press releases, public statements, media coverage, stage of product life cycle, product's price, competing and substitutable products' prices, seasonality effect, and product's previous sales. Besides controlling for such factors, such a study would need to control for selection bias, since a blogger may choose to write about a product for all of the other plausible reasons mentioned. Therefore, establishing causality between posts and product sales would be challenging. Likewise, it may be difficult to tease out the effect of employee blogging on the price of the company's stock. Though a lot of market measures are easily available for public companies, adding variables in a model may add variability in the model parameters and such a comprehensive study would need to work with many more variables. It would therefore be difficult to detect the slight effect of employee blogs on company stock price. In this study, we have access to aggregate readership of a blog at week level. Hence, we cannot identify the readers. If one could

identify the readers it would be interesting to see how negative posts attract new readers and retain old readers. It would also be interesting to see how the context in which posts of a blog are cited affects its readership. Further, our study needs to be replicated on other datasets to ensure generalizability of our findings. We have reported the results for a technology firm's blogs. Future research can study if the results are similar across industries.

For future research, it will be interesting to see how company-specific characteristics affect our framework, but such analysis will require readership data of employee blogs of different companies during the same time period. Another interesting direction would be to study how different textual characteristics of a post or blogger demographics affect the readership as well as the influence of a post. Such a study would help firms move closer to the optimal solution for their employees' blogs. For example, if the outcome of the study is that promoting a product of a firm's competitor has much higher influence on average than criticizing the firm's own product, and the firm wants to encourage slight negative content on employee blogs, it may ask its employees not to promote competitors' products but instead encourage them to discuss drawbacks of the firm's products.

References:

- Adamic, L., and N. Glance. 2005. The political blogosphere and the 2004 election: divided they blog. Paper read at International Conference on Knowledge Discovery and Data Mining (KDD), at Chicago, IL, USA.
- Alexander, J. C., Jr., and R. H. Mabry. 1994. Relative significance of journals, authors, and articles cited in financial research. *Journal of Finance* 49 (2):697.
- Arellano, M., and S. Bond. 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies* 58 (2):277-297.
- Bates, D. M., and D. G. Watts. 1988. *Nonlinear Regression Analysis and Its Applications*: Wiley.
- Battelle, J. 2005. *The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture*: Penguin Group.
- Baumgartner, H., and R. Pieters. 2003. The Structural Influence of Marketing Journals: A Citation Analysis of the Discipline and Its Subareas Over Time. *Journal of Marketing* 67 (2):123.
- Brian, S. 2005. Corporate Marketers Try Out Blogs. *Wall Street Journal*, B.8.
- Brin, S., and L. Page. 1998. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks & ISDN Systems* 30 (1-7):107.

- Byron, D., and S. Broback. 2006. *Publish and Prosper: Blogging for Your Business*: New Riders Press.
- Coffin, T. E. 1963. A Pioneering Experiment in Assessing Advertising Effectiveness. *Journal of Marketing* 27 (3):1-10.
- Crawford, K. 2005. Have a Blog, Lose Your Job? *CNN/Money*.
- Culnan, M. J. 1986. The Intellectual Development of Management Information Systems, 1972-1982: A Co-Citation Analysis. *Management Science* 32 (2):156-172.
- . 1987. Mapping the Intellectual Structure of MIS, 1980-1985: A Co-Citation Analysis. *MIS Quarterly* 11 (3):341.
- Davidson, R., and J. G. MacKinnon. 1993. *Estimation and Inference in Econometrics*: Oxford University Press.
- Drezner, D. W., and H. Farrell. 2004. The power and politics of blogs. Paper read at American Political Science Association.
- Eckelberry, A. *Yankee Group/Sunbelt 2006 Server Reliability Survey Results* Sunbelt Blog 2006 [cited. Available from <http://sunbeltblog.blogspot.com/2006/03/yankee-groupsunbelt-2006-server.html>].
- Edelmen, and Intelliseek. 2005. Talking From The Inside Out: The Rise of Employee Bloggers. In *New Communications and Word-of-Mouth Marketing*.
- Farrell, H., and D. Drezner. 2008. The power and politics of blogs. *Public Choice* 134 (1):15-30.
- Folkes, V. S. 1988. Recent Attribution Research in Consumer Behavior: A Review. *Journal of Consumer Research* 14 (4):548.
- Fotak, V. 2008. The Impact of Blog Recommendations on Security Prices and Trading Volumes. *Financial Management Association*.
- Fred, R. S. 1992. Origins of bibliometrics, citation indexing, and citation analysis: The neglected legal literature, 337-339.
- Furukawa, T., M. Ishizuka, Y. Matsuo, I. Ohmukai, and K. Uchiyama. 2007. Analyzing reading behavior by blog mining. *Association for the Advancement of Artificial Intelligence (AAAI)* 22:1353.
- Ghose, A., and P. G. Ipeirotis. 2010. Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics. *IEEE Transactions on Knowledge and Data Engineering (TKDE)* Forthcoming.
- Ghose, A., P. G. Ipeirotis, and A. Sundararajan. 2009. The dimensions of reputation in electronic markets: NYU.
- Gittelman, M., and B. Kogut. 2003. Does good science lead to valuable knowledge? Biotechnology firms and the evolutionary logic of citation patterns. *Management Science* 49 (4):366.
- Goldstone, A. 2006. MP3 Blogs: A Silver Bullet for the Music Industry or a Smoking Gun for Copyright Infringement? *SSRN eLibrary*.
- Hall, B. H., A. Jaffe, and M. Trajtenberg. 2005. Market value and patent citations. *The Rand Journal of Economics* 36 (1):16.

- Halley, S. 2003. A blogger in their midst. *Harvard Business Review* 81 (9):30.
- Harhoff, D., F. Narin, F. M. Scherer, and K. Vopel. 1999. Citation frequency and the value of patented inventions. *The Review of Economics and Statistics* 81 (3):511.
- Hausman, J. A. 1978. Specification tests in econometrics. *Econometrica* 46 (6):1251-1271.
- Headen, R. S., J. E. Klompmaker, and J. E. Teel, Jr. 1977. Predicting Audience Exposure to Spot TV Advertising Schedules. *Journal of Marketing Research* 14 (1):1-9.
- Heckman, J. J., and B. Singer. 1984. Econometric duration analysis. *Journal of Econometrics* 24 (1-2):63-132.
- Jaccard, J., and R. Turrisi. 2003. *Interaction Effects in Multiple Regression*. Edited by M. S. Lewis-Beck. Second ed: Saga Publications.
- Jones, E. E., D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, and B. Weiner. 1987. *Attribution: Perceiving the Causes of Behavior*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Kelley, H. H. 1972. Attribution in Social Interaction. In *Attribution: Perceiving the Causes of Behavior*, edited by E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, K. Vopel and B. Weiner. Morristown, NJ: General Learning Press, 1-26.
- Kelley, H. H., and J. L. Michela. 1980. Attribution Theory and Research. *Annual Review of Psychology* 31:457-501.
- Kirkpatrick, D. 2005. Why There's No Escaping the Blog. *Fortune*, January 10, 2005.
- Kosma, M. N. 1998. Measuring the Influence of Supreme Court Justices. *The Journal of Legal Studies* 27 (2):333-372.
- Landes, W. M., L. Lessig, and M. E. Solimine. 1998. Judicial Influence: A Citation Analysis of Federal Courts of Appeals Judges. *The Journal of Legal Studies* 27 (2):271-332.
- Leskovec, J., M. McGlohon, C. Faloutsos, N. Glance, and M. Hurst. 2007. Cascading Behavior in Large Blog Graphs. In *SIAM International Conference on Data Mining (SDM)*. Minnesota, MN, USA.
- Lu, Y., P. Tsaparas, A. Ntoulas, and L. Polanyi. 2010. Exploiting social context for review quality prediction. Paper read at Proceedings of the 19th international conference on World Wide Web (WWW), at Raleigh, North Carolina, USA.
- Massie, R., and K. Kurapati. *There's Blogs (and Feeds!) In Our Search!* Blog.ask.com, June 01, 2006 2006 [cited. Available from http://blog.ask.com/2006/06/theres_blogs_an.html].
- Medoff, M. H. 1996. A citation-based analysis of economists and economics programs. *American Economist* 40 (1):46.
- Michaelis, L., and M. L. Menten. 1913. Die Kinetik der Invertinwirkung. In *Biochem. Z.*, 333-369.
- Mishne, G., and N. Glance. 2006. Predicting Movie Sales from Blogger Sentiment. Paper read at AAAI 2006 Spring Symposium on Computational Approaches to Analysing Weblogs.
- Miura, A., and K. Yamashita. 2007. Psychological and social influences on blog writing: An online survey of blog authors in Japan. *Journal of Computer Mediated Communication* 12 (4):1452.

- Nardi, B. A., D. J. Schiano, M. Gumbrecht, and L. Swartz. 2004. Why we blog. *Association for Computing Machinery. Communications of the ACM* 47 (12):41.
- Nielsen. 2007. 'Word-of-Mouth' the most powerful selling tool: Nielsen Global Survey. In *Trust in Advertising*, 1-6.
- Oliver, R. 2006. Blogger in Chief. *Fortune* 154 (10):51.
- Onishi, H., and P. Manchanda. 2009. Marketing Activity, Blogging and Sales. *SSRN eLibrary*.
- Pedersen, S., and C. Macafee. 2007. Gender differences in British blogging. *Journal of Computer Mediated Communication* 12 (4):1472.
- Pinski, G., and F. Narin. 1976. Citation Influence for Journal Aggregates of Scientific Publications: Theory, with Application to the Literature of Physics. *Information Processing and Management*.
- Posner, R. A. 2000. An economic analysis of the use of citations in the law. *American Law and Economics Review* 2 (2):381-406.
- Ramos-Rodríguez, A.-R., and J. Ruíz-Navarro. 2004. Changes in the intellectual structure of strategic management research: a bibliometric study of the Strategic Management Journal, 1980-2000. *Strategic Management Journal* 25 (10):981.
- Schwartz, J. 2005. If You Want to Lead, Blog. *Harvard Business Review* 83 (11):30.
- Scoble, R. *Windows More Reliable Than RedHat Linux? Yes, says Yankee Group.* <http://scobleizer.wordpress.com> 2006 [cited. Available from <http://scobleizer.com/2006/06/06/windows-more-reliable-than-redhat-linux-yes-says-yankee-group/>].
- Scoble, R., and S. Israel. 2006. *Naked conversations: How Blogs are Changing the Way Businesses Talk with Customers*: John Wiley.
- Seber, G. A. F., and C. J. Wild. 2003. *Nonlinear Regression*: Wiley-Interscience.
- Shirky, C. 2003. Power Laws, Weblogs, and Inequality. In *Exposure: From Friction to Freedom*: Aula.
- Stigler, G. J., and C. Friedland. 1975. The Citation Practices of Doctorates in Economics. *The Journal of Political Economy* 83 (3):477.
- Stock, J. H., J. H. Wright, and M. Yogo. 2002. A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics* 20 (4):518-529.
- Tahai, A., and M. J. Meyer. 1999. A revealed preference study of management journals' direct influences. *Strategic Management Journal* 20 (3):279.
- Trajtenberg, M. 1990. A Penny for Your Quotes: Patent Citations and the Value of Innovations. *The Rand Journal of Economics* 21 (1):172.
- Weil, D. 2010. *The Corporate Blogging Book: Absolutely Everything You Need to Know to Get It Right*. Updated 1st edition ed: WordBiz.com, Inc.
- Weiner, B. 2000. Attributional thoughts about consumer behavior. *Journal of Consumer Research* 27 (3):382.

Wood, W., and A. H. Eagly. 1981. Stages in the analysis of persuasive messages: The role of causal attributions and message comprehension. *Journal of Personality and Social Psychology* 40 (2):246-259.

Wooldridge, J. M. 2001. *Econometric Analysis of Cross Section and Panel Data*. 1 ed: The MIT Press.

Wright, J. 2005. *Blog Marketing: The Revolutionary New Way to Increase Sales, Build Your Brand, and Get Exceptional Results*. 1st ed: McGraw-Hill.

Yardi, S., S. A. Golder, and M. J. Brzozowski. 2009. Blogging at work and the corporate attention economy. In *Proceedings of the 27th international conference on Human factors in computing systems*. Boston, MA, USA: ACM.

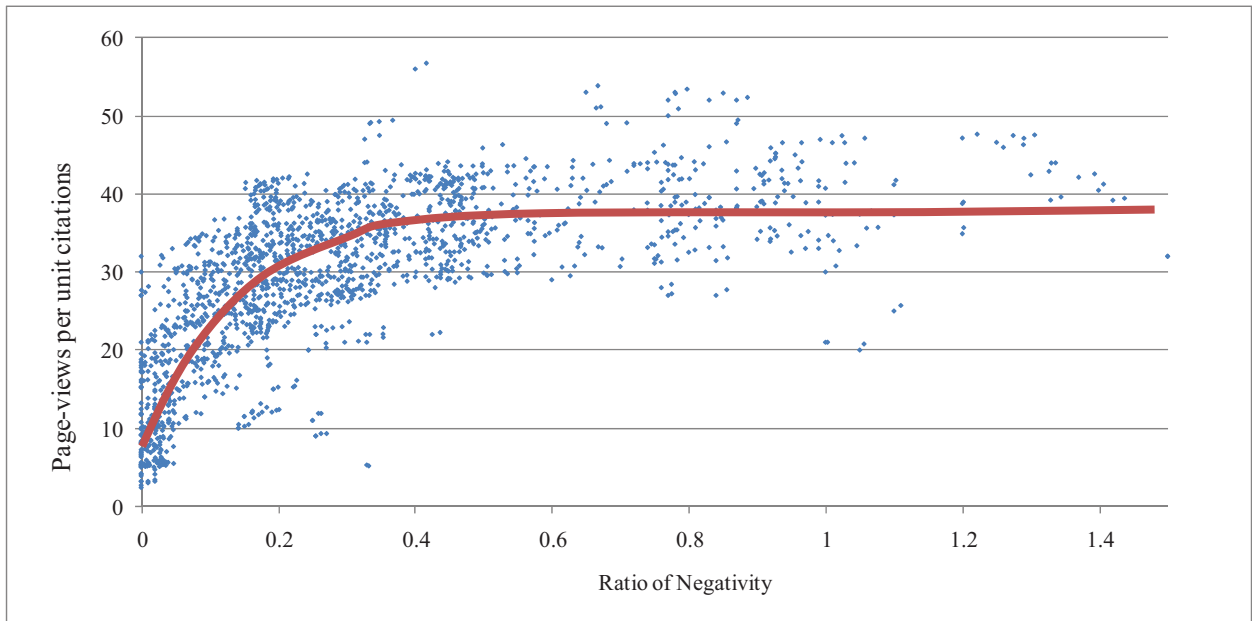


Figure 1: Page-views per unit citations vs. ratio of negativity

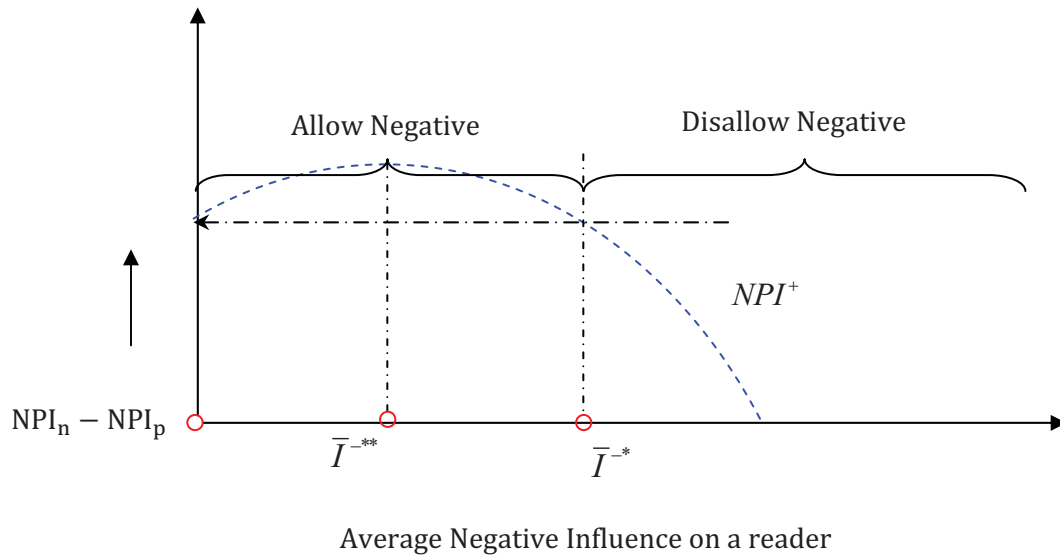


Figure 2: Decisions corresponding to average negative influence on a reader

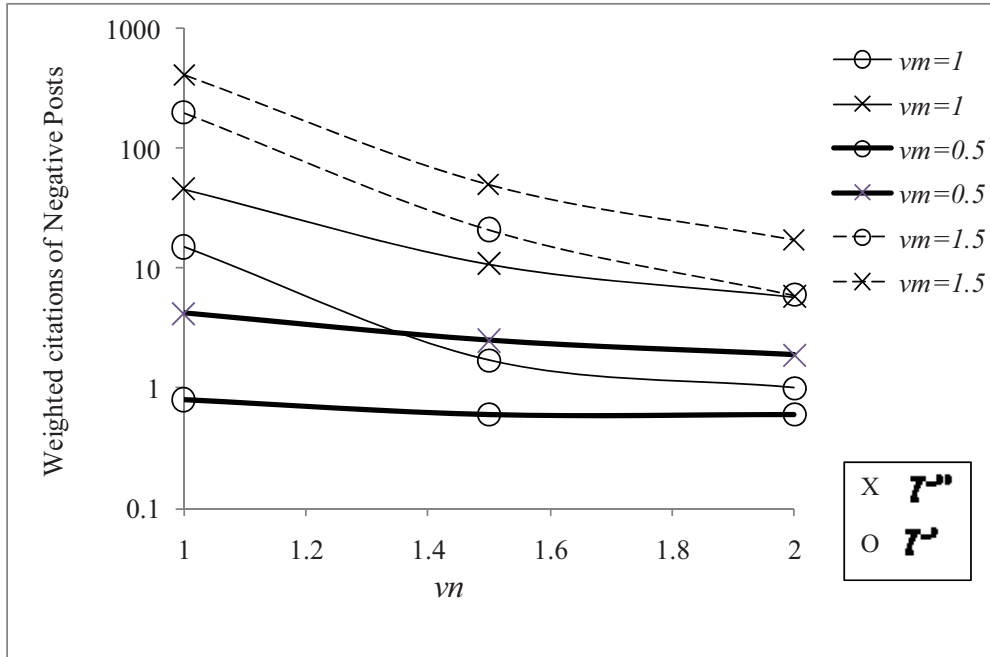


Figure 3a: Optimal weighted citations of negative posts vs vn .

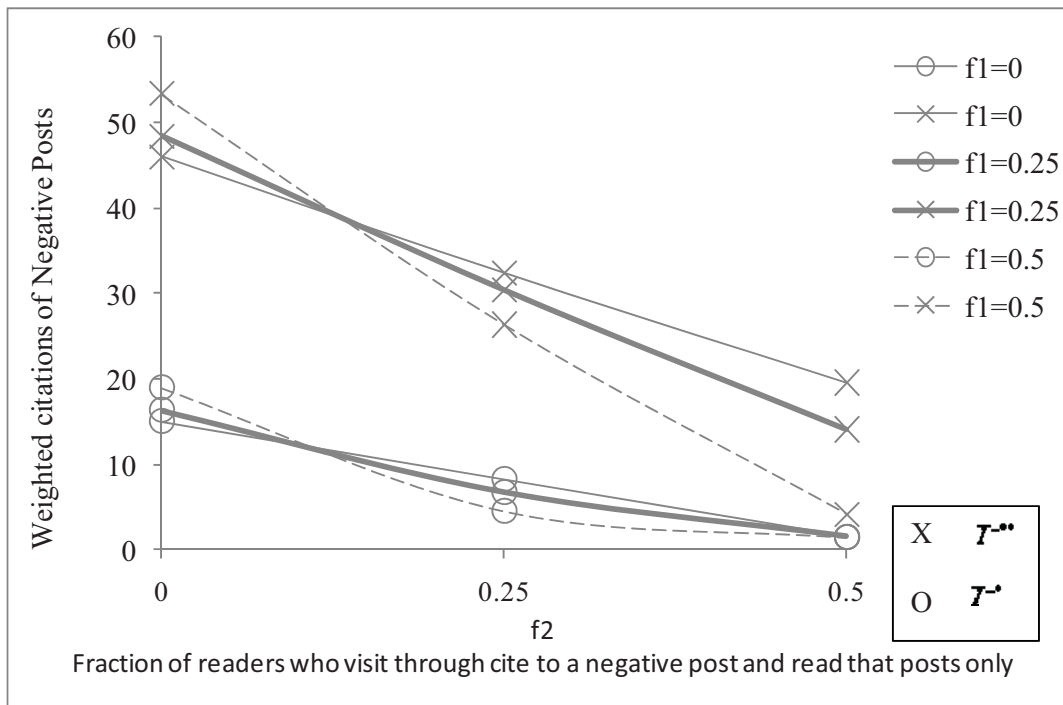


Figure 3b: Optimal weighted citations of negative posts vs $f2$.

(Note: $f1$ is the fraction of readers who visited the blog through cite to a positive post and read that post only)

Table 1: Definitions and Interpretations of Variables

Variable	Definition	Interpretation
Page-views ($P_{i,t}$)	Number of views of blog i in period t	High page-views indicates higher number of readers
Page-views Lag ($P_{i,t-1}$)	Number of views of blog i in period t-1	High page-views lag indicates higher popularity of a blog
Weighted Citation of a blog ($W_{i,t}$)	Number of Weighted citations received by blog i for posts displayed in period t.	Higher weighted citations indicate that the blog posts are considered important by the blogging community.
Weighted Citation of positive posts of a blog ($W_{i,t}$)	Number of Weighted citations received by blog i for positive posts displayed in period t.	Higher weighted citations indicate that the blog posts that say positive things about the blogger's firm are considered important by the blogging community.
Weighted Citation of negative posts of a blog ($W_{i,t}$)	Number of Weighted citations received by blog i for negative posts displayed in period t.	Higher weighted citations indicate that the blog posts that say negative things about the blogger's firm are considered important by the blogging community.
Influence of a blog on a reader ($I_{i,t}$)	Extent to which blog i affects views of a reader in period t	High influence of a blog indicates that a blog affects a reader's views more
Positive influence of a blog on a reader ($I_{i,t}^+$)	Total positive effect of blog i on the views of a reader in period t towards the employer of the blogger	High positive influence of a blog indicates that a blog has a high total positive effect on a reader's views toward the employer of the blogger
Negative influence of a blog on a reader ($I_{i,t}^-$)	Total negative effect of blog i on the views of a reader in period t towards the employer of the blogger	High negative influence of a blog indicates that a blog has a high total negative effect on a reader's views toward the employer of the blogger
Net positive influence of a blog on a reader ($I_{i,t}^+ - I_{i,t}^-$)	Total overall positive effect of blog i on the views of a reader in period t towards the employer of the blogger	High net positive influence of a blog on a reader indicates that a blog has a high overall positive effect on a reader's views toward the employer of the blogger
Net positive influence of a blog on readers ($NPI_{i,t}^{++}$)	Total overall positive effect of blog i on the views of readers in period t towards the employer of the blogger	High net positive influence of a blog on readers indicates that a blog has a high overall positive effect on readers' views toward the employer of the blogger
Ratio of negativity ($R_{i,t}$)	Sum of weighted citations of negative posts / Sum of weighted citations of positive posts of blog i in period t	High ratio of negativity indicates higher proportion of negative content on a blog
Subscribers ($S_{i,t}$)	XML feed subscribers of blog i in period t	More subscribers indicate higher popularity of a blog

Table 2: Descriptive Statistics

Variable	Mean	Min.	Max.	Std. Dev.
Page-views ($P_{i,t}$)	2402.77	96	4587	820.34
Citations Received by blog ($W_{i,t}$)	84.21	0	130.23	55.74
Citations received by Positive Posts ($W_{i,t}^+$)	56.56	0	98.66	36.53
Citations received by Negative Posts ($W_{i,t}^-$)	12.31	0	85.74	39.31
Citations received by Neutral Posts ($W_{i,t}^0$)	2.34	0	26.52	9.46
Ratio of Negativity ($R_{i,t}$)	0.15	0	1.5	0.39
No. of Positive Posts	7.24	4	15	3.14
No. of Negative Posts	0.97	0	6	1.08
No. of Neutral Posts	6.79	0	10	2.46
No. of Subscriber ($S_{i,t}$)	97.11	16	295	89.94

Table 3: Effect of Negative Posts on Readership

Parameter	Pooled OLS Estimates
$P_{i,t-1}$	0.165*** (0.05)
$W_{i,t}$	18.467*** (7.64)
$R_{i,t}$	15.526 (15.35)
$W_{i,t}R_{i,t}$	15.757*** (4.93)
$S_{i,t}$	3.757 (4.98)
Constant	32.291*** (14.21)
Adj. R^2	46.149%
N	2,321
$Pr>F$	0.000
RMSE	796.139

***= $p < 0.01$, **= $p < 0.05$, *= $p < 0.1$, std. errors in parenthesis

Table 4: Effect of Negative Posts on Readership [Posts separated by type]

Parameter	Pooled OLS Estimates
$P_{i,t-1}$	0.167*** (0.04)
$W_{i,t}^+$	18.491*** (5.35)
$W_{i,t}^-$	34.457*** (9.61)
$W_{i,t}^0$	15.478*** (4.33)
$R_{i,t}$	15.264 (16.35)
$W_{i,t}^0 R_{i,t}$	14.567*** (3.93)
$W_{i,t}^- R_{i,t}$	31.526*** (4.67)
$S_{i,t}$	3.691 (5.02)
Constant	31.163*** (12.25)
Adj. R^2	51.312%
N	2,321
$Pr>F$	0.000
RMSE	836.195

***= $p<0.01$, **= $p<0.05$, *= $p<0.1$, std. errors in parenthesis

Table 5: Effect of Negative Posts on Readership – Michaelis Menten Model

Parameter	Maximum Likelihood Estimates
$P_{i,t-1}$	0.129*** (0.01)
$W_{i,t}^+$	14.334** (2.99)
$W_{i,t}^-$	29.718*** (4.47)
$W_{i,t}^0$	11.371*** (1.06)
$W_{i,t}^+ R_{i,t} / (R_{i,t} + \beta_4)$	13.546*** (2.08)
$W_{i,t}^- R_{i,t} / (R_{i,t} + \beta_4)$	33.293*** (3.37)
$W_{i,t}^0 R_{i,t} / (R_{i,t} + \beta_4)$	9.952 * (4.89)
β_4	0.307*** (0.02)
$R_{i,t}$	12.116 (13.95)
$S_{i,t}$	3.054 (4.12)
Constant	34.009*** (8.22)
N	2321
LogLikelihood	-291109.32

***= $p<0.01$, **= $p<0.05$, *= $p<0.1$, std. errors in parenthesis

Table 6: Effect of Negative Posts on Readership

Parameter	Blog Random Effects AR1 (IV) Estimates	Blog Random Effects AR1 (IV) Estimates	Blog First Difference AR1 (IV) Estimates
$P_{i,t-1}$	0.119*** (0.01)	0.122*** (0.01)	0.127*** (0.05)
$W_{i,t}^+$	16.212*** (2.19)	16.917*** (2.18)	15.879*** (3.94)
$W_{i,t}^-$	36.011*** (3.72)	34.259*** (3.74)	38.191*** (5.35)
$W_{i,t}^0$	7.207*** (1.97)	7.618*** (1.97)	7.494*** (2.29)
$W_{i,t}^+ R_{i,t} / (R_{i,t} + \beta_4)$	10.463*** (3.11)	9.441*** (3.13)	10.415*** (4.01)
$W_{i,t}^- R_{i,t} / (R_{i,t} + \beta_4)$	28.134*** (4.72)	28.834*** (4.61)	27.356*** (6.98)
$W_{i,t}^0 R_{i,t} / (R_{i,t} + \beta_4)$	5.151 ** (2.19)	5.104 ** (2.11)	7.159 ** (3.49)
β_4	0.497*** (0.02)	0.489*** (0.03)	0.479*** (0.07)
$R_{i,t}$	12.121 (14.02)	15.935 (18.87)	18.291 (24.54)
$S_{i,t}$	3.031 (4.07)	3.919 (4.76)	5.068 (5.13)
Constant	35.699*** (8.36)	35.385*** (8.22)	--
ρ	0.07***(0.001)	0.07***(0.001)	0.07***(0.001)
Control Variables	--	Time	--
N	2321	2321	2321
Log Likelihood	-284619.52	-284614.92	-246106.48

***= $p < 0.01$, **= $p < 0.05$, *= $p < 0.1$, robust std. errors in parenthesis ; Likelihood Ratio Test Results (Time dummies are insignificant); Because a constant term is estimated, the mean for blog random effects is set to zero. Blog random effects are estimated by a non parametric estimation (Heckman and Singer 1985).

Table 7: Models Used for Comparison

$$1) P_{i,t} = \beta_1 P_{i,t-1} + \beta_2 W_{i,t}^+ + \beta_3 W_{i,t}^- + \beta_4 W_{i,t}^0 + \left(\beta_5 W_{i,t}^+ + \beta_6 W_{i,t}^- + \beta_7 W_{i,t}^0 \right) \frac{R_{i,t}}{R_{i,t} + \beta_8} + \beta_9 R_{i,t} + \beta_{10} S_{i,t} + \beta_{11i} + \gamma_{i,t}$$

$$2) P_{i,t} = \beta_1 P_{i,t-1} + \beta_2 W_{i,t}^+ + \beta_3 W_{i,t}^- + \beta_4 W_{i,t}^0 + \left(\beta_5 W_{i,t}^+ + \beta_6 W_{i,t}^- + \beta_7 W_{i,t}^0 \right) \exp(\beta_8 R_{i,t}) + \beta_9 R_{i,t} + \beta_{10} S_{i,t} + \beta_{11i} + \gamma_{i,t}$$

$$3) P_{i,t} = \beta_1 P_{i,t-1} + \beta_2 W_{i,t}^+ + \beta_3 W_{i,t}^- + \beta_4 W_{i,t}^0 + \left(\beta_5 W_{i,t}^+ + \beta_6 W_{i,t}^- + \beta_7 W_{i,t}^0 \right) \frac{R_{i,t}}{R_{i,t} + \beta_8 + \beta_{12} R_{i,t}^2} + \beta_9 R_{i,t} + \beta_{10} S_{i,t} + \beta_{11i} + \gamma_{i,t}$$

$$4) P_{i,t} = \beta_1 P_{i,t-1} + \beta_2 W_{i,t}^+ + \beta_3 W_{i,t}^- + \beta_4 W_{i,t}^0 + \beta_6 W_{i,t}^- R_{i,t} + \beta_7 W_{i,t}^0 R_{i,t} + \beta_{13} W_{i,t}^- R_{i,t}^2 + \beta_{14} W_{i,t}^0 R_{i,t}^2 + \beta_9 R_{i,t} + \beta_{10} S_{i,t} + \beta_{11i} + \gamma_{i,t}$$

$$5) P_{i,t} = \beta_1 P_{i,t-1} + \beta_2 W_{i,t}^+ + \beta_3 W_{i,t}^- + \beta_4 W_{i,t}^0 + \beta_6 W_{i,t}^- R_{i,t} + \beta_7 W_{i,t}^0 R_{i,t} + \beta_9 R_{i,t} + \beta_{10} S_{i,t} + \beta_{11i} + \gamma_{i,t}$$

$$\text{In all models: } Y_{i,t} = \eta Y_{i,t-1} + \epsilon_{i,t}, \epsilon_{i,t} \sim N(0, \sigma^2)$$

(Ceteris Paribus) In all the models with the increase in Negative Influence, Readership initially increase.

Table 8: Comparison of Potential Models

Models		Exponential & after a point stabilize		Exponentially & after a point fall		Linearly
		1	2	3	4	5
<i>IV Estimates; ARI Correlation in Errors; Blog Random Effects</i>	β_1	0.119*** (0.01)	0.125*** (0.03)	0.125*** (0.02)	0.145*** (0.05)	0.147*** (0.04)
	β_2	16.212*** (2.19)	18.925*** (3.26)	15.927*** (2.23)	18.991*** (2.19)	18.473*** (3.31)
	β_3	36.011*** (3.72)	39.997*** (4.82)	34.819*** (3.69)	32.457*** (3.72)	34.457*** (4.61)
	β_4	7.207*** (1.97)	9.975*** (2.38)	7.288*** (2.08)	15.478*** (1.97)	15.478*** (1.33)
	β_5	10.463*** (3.11)	-1.383* (0.68)	9.743*** (3.19)		
	β_6	28.134*** (4.72)	-2.918*** (1.39)	26.988*** (5.24)	32.526*** (4.72)	14.567*** (1.93)
	β_7	5.151 ** (2.19)	-0.592 (1.14)	5.216 ** (2.11)	15.567 ** (2.19)	31.526*** (3.67)
	β_8	0.497*** (0.02)	-6.105*** (1.22)	0.423*** (0.03)		
	β_9	12.121 (14.02)	15.947 (14.25)	12.983 (13.99)	13.046 (14.28)	15.264 (11.15)
	β_{10}	3.031 (4.07)	3.913 (4.37)	3.246 (4.18)	3.231 (4.11)	3.691 (3.41)
	β_{12}			0.098(1.21)		
	β_{13}				- 13.8115*** (1.74)	
	β_{14}				-7.187* (2.32)	
	β_{11}	35.699*** (8.36)	39.982*** (8.89)	35.074*** (8.42)	37.504*** (8.31)	32.032*** (8.21)
	ρ	0.07*** (0.001)	0.07*** (0.002)	0.07*** (0.001)	0.07*** (0.001)	0.07*** (0.002)
<i>Log Likelihood</i>		- 284619.52	- 292170.35	-284618.93	-294029.70	-297818.54
<i>Training</i>	<i>BIC</i>	569378.54	584480.20	579287.11	588198.90	595761.08
<i>Training</i>	<i>MAD</i>	614.324	645.601	635.081	704.189	754.098
<i>Testing</i>	<i>MAD</i>	819.209	856.39	855.35	890.36	899.19

***= $p < 0.01$, **= $p < 0.05$, *= $p < 0.1$, robust std. errors in parenthesis; Because a constant term is estimated, the mean for blog random effects is set to zero. Blog random effects are estimated by a non parametric estimation (Heckman and Singer 1984).

Note that main effect of Ratio of Negativity is insignificant in all models ($p < 0.25$)

Table 9: Effect of Negative Posts on Citations

Parameters	Blog Random Effects Negative Binomial Model	Blog Random Effects Negative Binomial Model with Interactions
$P_{i,t-1}$	2.441***(1.14)	2.362**(1.16)
$Post^+$	2.893**(1.41)	2.325*(1.19)
$Post^-$	45.271***(8.49)	43.983***(8.78)
$R_{i,t}$	6.13***(1.32)	2.896***(1.091)
$S_{i,t}$	0.21(0.29)	0.17(0.31)
$Post^-XR_{i,t}$		2.562***(0.97)
$Post^0XR_{i,t}$		3.727**(1.87)
$Post^+XR_{i,t}$		7.81***(2.50)
<i>Constant</i>	-1.41***(0.43)	-1.76***(0.38)
<i>Dispersion</i>	0.009***(0.001)	0.009***(0.001)
<i>N</i>	2743	2743
<i>Log Likelihood</i>	-18905.32	-18605.43

***= $p < 0.01$, **= $p < 0.05$, *= $p < 0.1$, robust std. errors in parenthesis; Because a constant term is estimated, the mean for blog random effects is set to zero. Blog random effects are estimated by a non parametric estimation (Heckman and Singer 1984).

Online Supplement to Blog, Blogger, and the Firm: Can Negative Posts by Employees Lead to Positive Outcomes?

APPENDIX A: MEASURING THE WEIGHTED CITATIONS OF A POST

Consider the citations to a post as shown below.

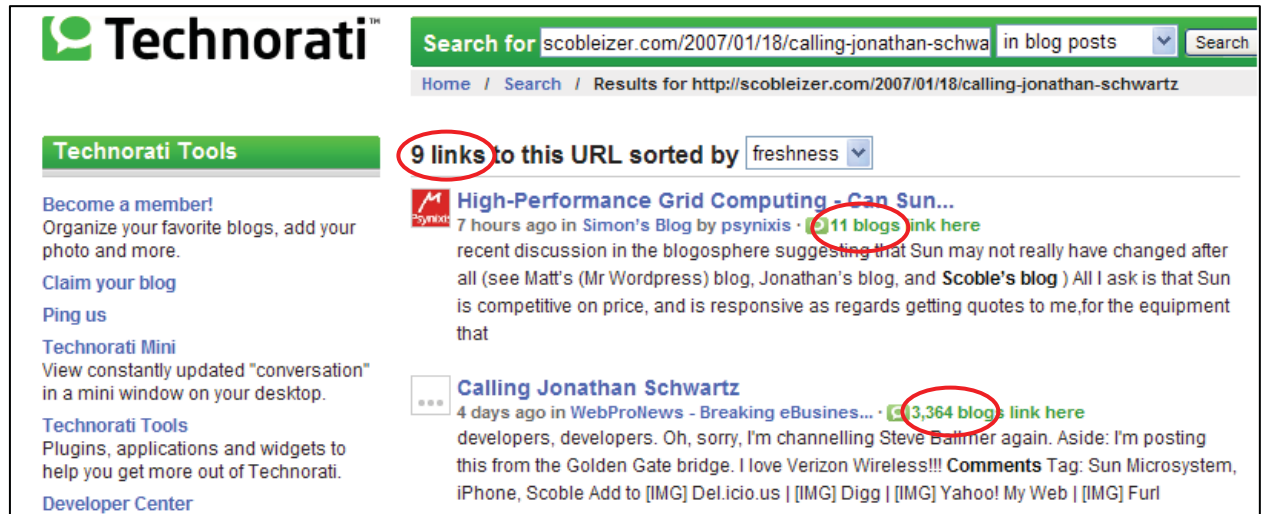


Figure A-1: Snapshot of the blog search engine, Technorati

Influence of a post, $S \equiv \# \text{ of citations} + \sum \ln(\# \text{ of } 2^{\text{nd}} \text{ level citations})$

For example, influence of the above post, $S \equiv 9 + \ln(11) + \ln(3364) + \dots$

The literature suggests two ways of accounting for second order citations: exponential smooth or logarithmic smoothing (Shirky 2003; Drezner and Farrell 2004). We have taken the log of second-level citations. Our results are robust regarding the choice of method for measuring weighted citations to a post. We have also done the analysis without considering second-level citations. The results are the same, except that the coefficients are larger without considering second-level citations than those with these citations.

APPENDIX B: PROOF OF PROPOSITION

The proposition states that there exists a cutoff value of the average negative influence \bar{I}^{-*} such that (a) for all values of the average negative influence on a reader $\bar{I}^{-} \in (0, \bar{I}^{-*})$, the decision of allowing negative posts creates more average net positive influence on readers than the decision of no negative posting, whereas (b) for $\bar{I}^{-} > \bar{I}^{-*}$, the decision of no negative posting offers a higher average net positive influence on readers than the decision of allowing negative posts. To prove this proposition, we need to prove that $NPI_n - NPI_p$ is positive and concave ($\frac{\partial^2(NPI_n - NPI_p)}{\partial^2(\bar{I}^{-})} < 0$). Note that NPI_p is always positive. Hence if we prove that the difference between the two NPI is concave it is a sufficient to prove the proposition.

For simplicity we would assume that $v_n = v_m = 1$. Once we assume this Weighted citations of a blog is exactly equal to its influence. Note that if $v_n > v_m$ would just reduce the space under which allowing negative posts is optimal. If $v_m > v_n$ it would increase the space under which allowing negative posts is optimal.

For Proof simplification, we would just work with the reduced model where

$$\bar{P}_n = \frac{1}{1 - \beta_1} \left(\beta_9 + \beta_2 \bar{W}_n + (\beta_5 \bar{W}_n) \left(\frac{\bar{W}_n}{\bar{W}_n + \beta_8 \bar{W}_n^+} \right) \right)$$

$$\bar{P}_p = \frac{1}{1 - \beta_1} \left(\beta_9 + \beta_2 \bar{W}_p + (\beta_5 \bar{W}_p) \left(\frac{\bar{W}_p}{\bar{W}_p + \beta_8 \bar{W}_p^+} \right) \right)$$

Note that this assumption is quite conservative as it reduces the effect of negative post in increasing page views to as much as that by positive or neutral post. This assumption biases NPI_n towards zero. Under these assumptions:

$$\frac{\partial^2(NPI_n - NPI_p)}{\partial \bar{I}^{-2}} = - \frac{2(\beta_2 \bar{I}^{-3} + 3\beta_2 \beta_4 \bar{I}^{-2} \bar{I}^+ + 3\beta_2 \beta_4^2 \bar{I}^{-} \bar{I}^{+2} + \beta_2 \beta_4^3 \bar{I}^{+3} + \beta_3 \bar{I}^{-3} + 3\beta_3 \beta_4 \bar{I}^{-2} \bar{I}^+ + 3\beta_3 \beta_4^2 \bar{I}^{-} \bar{I}^{+2} + \beta_3 \beta_4^2 \bar{I}^0 \bar{I}^{+2} + \beta_3 \beta_4 \bar{I}^{+3} + \beta_3 \beta_4 \bar{I}^{+2} \bar{I}^0)}{(\bar{I}^{-} + \beta_4 \bar{I}^+)^3} \quad (1)$$

In the above expression, note that all values in the numerator and denominator are positive, and the expression overall has a negative sign.

$NPI_n = NPI_p$ under two conditions: (1) when $\bar{I}^- = 0$, because when there are no negative posts, NPI^{-+} and NPI^+ are identical, and (2) $\bar{I}^- = \bar{I}^{-*}$, where \bar{I}^{-*} solves $NPI_n = NPI_p$. This expression is quadratic in \bar{I}^- , so solving for \bar{I}^{-*} and ignoring the negative root yields:

$$\frac{\sqrt{\beta_2^2 \bar{I}^{0^2} + 2\beta_2\beta_3 \bar{I}^{0^2} - 2\beta_2^2\beta_4 \bar{I}^0 \bar{I}^+ + \beta_3^2 \bar{I}^{0^2} - 2\beta_2\beta_3\beta_4 \bar{I}^0 + \beta_2^2\beta_4^2 \bar{I}^{+2} + 4\beta_2\beta_3 \bar{I}^{+2} + 4\beta_2\beta_3 \bar{I}^0 \bar{I}^+ + 4\beta_3^2 \bar{I}^{+2} + 4\beta_3^2 \bar{I}^0 \bar{I}^+} - (\beta_2 \bar{I}^0 + \beta_3 \bar{I}^0 + \beta_2\beta_4 \bar{I}^+)}{2(\beta_2 + \beta_3)}$$

We find that $\beta_3 > \beta_2\beta_4$ is a sufficient condition for \bar{I}^{-*} to exist and be positive. From our empirical analysis, the data satisfies the condition $\beta_3 > \beta_2\beta_4$. Let

$$a = \frac{\sqrt{\beta_2^2 \bar{I}^{0^2} + 2\beta_2\beta_3 \bar{I}^{0^2} - 2\beta_2^2\beta_4 \bar{I}^0 \bar{I}^+ + \beta_3^2 \bar{I}^{0^2} - 2\beta_2\beta_3\beta_4 \bar{I}^0 + \beta_2^2\beta_4^2 \bar{I}^{+2} + 4\beta_2\beta_3 \bar{I}^{+2} + 4\beta_2\beta_3 \bar{I}^0 \bar{I}^+ + 4\beta_3^2 \bar{I}^{+2} + 4\beta_3^2 \bar{I}^0 \bar{I}^+}}$$

$$\text{and } b = \beta_1 \bar{I}^0 + \beta_2 \bar{I}^0 + \beta_1\beta_3 \bar{I}^+.$$

When $\beta_3 > \beta_2\beta_4$, it can be seen that:

$$\begin{aligned} a &= \frac{\sqrt{\beta_2^2 \bar{I}^{0^2} + 2\beta_2\beta_3 \bar{I}^{0^2} - 2\beta_2^2\beta_4 \bar{I}^0 \bar{I}^+ + \beta_3^2 \bar{I}^{0^2} - 2\beta_2\beta_3\beta_4 \bar{I}^0 + \beta_2^2\beta_4^2 \bar{I}^{+2} + 4\beta_2\beta_3 \bar{I}^{+2} + 4\beta_2\beta_3 \bar{I}^0 \bar{I}^+ + 4\beta_3^2 \bar{I}^{+2} + 4\beta_3^2 \bar{I}^0 \bar{I}^+}}{\sqrt{(4\beta_2\beta_3 + 4\beta_3^2 + \beta_2^2\beta_4^2) \bar{I}^{+2} + (-2\beta_2^2\beta_4 \bar{I}^0 + 4\beta_2\beta_3 \bar{I}^0 - 2\beta_2\beta_3\beta_4 \bar{I}^0 + 4\beta_3^2 \bar{I}^0) \bar{I}^+ + \beta_2^2 \bar{I}^{0^2} + 2\beta_2\beta_3 \bar{I}^{0^2} + \beta_3^2 \bar{I}^{0^2}}} \\ &> \sqrt{(4\beta_2\beta_3 + 4\beta_3^2 + \beta_2^2\beta_4^2) \bar{I}^{+2} + (2\beta_2^2\beta_4 \bar{I}^0 + 2\beta_2\beta_3\beta_4 \bar{I}^0) \bar{I}^+ + \beta_2^2 \bar{I}^{0^2} + 2\beta_2\beta_3 \bar{I}^{0^2} + \beta_3^2 \bar{I}^{0^2}} > 0 \end{aligned}$$

This proves that a exists and is positive. Moreover, b is also positive. Now, we show that $a^2 - b^2$ is positive:

$$a^2 - b^2 = 4(\beta_2 + \beta_3) \bar{I}^+ (\beta_3 \bar{I}^+ + (\beta_3 - \beta_2\beta_4) \bar{I}^0) > 0$$

Algebraically, if $a^2 - b^2 > 0$ and $a + b > 0$, then $a - b > 0$. Hence, $\bar{I}^{-*} > 0$ (Proved).