The Relationship Between
Online Communications and Firm Value

Leigh McAlister

Garrett Sonnier

Tom Shiveley

July 2009

The authors would like to thank an anonymous sponsor company and the McCombs School's Center for Customer Insight for supporting this research.

ABSTRACT:

The visible trace of online communications has given rise to research on their effect on firm outcomes. The literature has established a link between online communication about a product and the product's sales and price performance. On the assumption that financial markets understand this link, we conjecture financial markets consider the amount of online communication, or chatter, about a firm to be an indication of the firm's performance in the marketplace. Our results confirm this conjecture. The relationship between stock returns and chatter are robust to alternative specifications of the model and to alternative measures of stock returns. We also investigate the issues of reverse causality and omitted variable bias due to omitted sales, profits or new product launches driving a spurious relationship between stock returns and chatter. The data are not consistent with any of these alternative explanations for our results.

July 2009

Introduction

As recently as ten years ago, most two-way product related communication took place in person or over the phone. Today, billions of people have product related conversations via Internet-based media like blogs, chat rooms and on-line forums (Godes et al. 2005). Second generation Internet sites, especially blogs and discussion forums, have facilitated information exchange and the diffusion of user generated content. From the perspective of consumers, product related communications are now scalable. Rather than one-to-one or one-to-few, such communications are one-to-many, with interpersonal communication feasible between the many participants. Interpersonal communication is often an influential source of product information for consumers (Godes and Mayzlin 2004). Online communications are also, for the most part, observable. From the perspective of researchers, observablity facilitates measurement of online communications, which has enabled marketing scholars to study their effect on various outcomes.

Anecdotal evidence suggests we should expect a link between online communication and sales. For example, BIGresearch.com (2005) reports that interpersonal communication (i.e., word of mouth) is the most influential source among various media in making a purchase. A stream of research on online communication establishes its impact on sales and prices. Godes and Mayzlin (2004) find that the dispersion of online conversations across different television based newsgroups explains variation in television ratings. Using web crawler technology to capture all online mentions of a product, Shin, Hanssens and Gajula (2008) show that positive online communication about a product is a leading indicator of retail prices. Focusing more precisely on online reviews, Chevalier and Mayzlin (2006) and Senecal and Nantel (2004) show that online customer book reviews are related to online sales. Liu (2006) and Dellarocas and Zhang and Awhad (2007) demonstrate that online customer movie reviews are related to movie revenues. Comparing reviews (both online and offline) with the simple volume of online communications, Dhar and Chang (2007) find that the latter is a better predictor of sales performance than the former.

The marketing and finance literatures have also considered the link between online and offline communications and firm stock returns. Using data from the airline industry, Luo (2007, 2008) shows that customer complaints are related to switching by future consumers, reducing revenue and, consequently, stock returns. Wysocki (1999) shows that increased postings on a firm's online stock message board are associated with financial news about the firm. Antweiler and Frank (2004) and Tumarkin and Whitelaw (2001) find that the volume of comments about a firm on an online stock message board is related to the variability of that firm's stock price. Das and Chen (2007) demonstrate that the volume and valence of comments about stocks in an index are related to the market price of the index.

Based on the accumulating literature, it is our hypothesis that if one looks beyond financially-oriented stock message boards and beyond the online communications or reviews specific to a single product to consider all of the online discussion about a firm on the Internet, the overall volume of online communication should be related to firm financial performance (e.g., firm revenues or firm profits). Thus, market participants may perceive online communications as reflective of information likely to affect cash flows, suggesting a mechanism by which online communications could impact firm stock returns. In this paper, we investigate the value relevance of online communications. The concept of value relevance refers to the question of whether a data series provides explanatory power beyond that provided by standard methods of explaining stock price movements (Mizik and Jacobson 2008).

Following Dhar and Chang (2007) we refer to overall online communications about a firm as “chatter”. There are several reasons to expect that chatter has value relevance. As discussed, the literature has shown online communication (including but not limited to online reviews) is related to product performance (i.e., sales or profits). Second, online communication is observable by the market. Investors may observe online communications on their own. In addition, aggregate information about online communication is available to the market via both private sources (e.g., firms such as Mango Analytics and Visible Technologies) and public sources (e.g., websites such as Technorati). Lastly (and importantly), the information is available at finer intervals than firm quarterly reports. Thus, financial market participants could exploit this intra-quarter insight into a firm’s performance.

Using a novel dataset provided to us by a technical firm, we look for a relationship between weekly stock returns and shocks to weekly chatter. The chatter measures provided to us cover a wide variety of topics and are classified as positive, negative, and neutral. Our technical firm is ideally suited to exploring the relationship between chatter and stock return. Technical products have become increasingly complex, driving consumers to the Internet for information before a purchase (Godes et al. 2005). Importantly, consumers also rely on the Internet for help in using technical products post-purchase. This is especially true for new technology products (Godes et al. 2005). We also might expect that differentiated technological products command higher margins. Using the firm's internal accounting data, we find significant correlation between weekly chatter and weekly gross profit margin (ρ =0.41). This suggests to the extent that they do, investors may very well be justified in treating chatter as a signal of unobserved intra-quarter firm performance.

To test the value relevance of chatter, we employ a model of weekly stock return based on the capital asset pricing model. We augment the usual independent variables in the model with dummy variables for publicly reported quarterly financial reports and measures of unexpected shocks to weekly chatter. We estimate models that aggregate chatter into a total measure as well as models that consider positive, negative and neutral chatter. We find that shocks to total weekly chatter are associated with stock returns. For the models that consider positive, negative, and neutral chatter, we find that shocks to neutral chatter are associated with stock returns. These findings are robust to alternative measures of stock return (e.g., cumulative abnormal returns and compounded abnormal returns). We consider issues of reverse causality and omitted variable bias. The data do not appear to be consistent with reverse causality. Lastly, the data do not support sales, profits or new product launches as being omitted variables causing a spurious association between chatter and return.

The remainder of our paper is structured as follows. In the next section, we present and discuss some pertinent features of our chatter data. We then present the models used to test the value relevance of the chatter data, followed by the results and a variety of robustness checks. The final section summarizes and concludes.

Data

Our chatter data are similar to Dhar and Chang (2007) and Shin, Hanssens and Gajula (2008). Both of these papers go beyond a single Internet forum to consider all online mentions of a particular product on the Internet. Our chatter data were collected by Visible Technologies, using their proprietary crawler technology that canvasses the Internet for mentions of the firm and/or the firm's products and/or services. The data consist of the count of online posts on various Internet sites (posts), the count of unique sites on which the posts appeared (sites), and the count of unique authors that generate the posts (authors). We observe 52 weeks of data from 2007. Figure 1 presents the mix of topics captured by the data. Management of the technical firm provided Visible with guidelines on how to assign the comments to topics. Brand and image topics (34%) and product topics (30%) are the most discussed. Notably, only about 1% of the topics are categorized as financial. Consistent with the notion that much of the chatter is related to making the company's technologically sophisticated products functional, about 29% of the topics relate to solving customer problems (customer service, outreach, tech support, etc.). For the remainder of the paper, we will aggregate the data across these topics into summary measures of posts, sites, and authors.

For each of the three categories (posts, sites, and authors) the data are classified by a proprietary algorithm as positive, negative, or neutral. Table 1 presents some descriptive statistics for the weekly counts. The neutral category accounts for the majority of observations and it also exhibits the most variability across time. Table 2 presents the correlations between these variables. Within each category of positive, negative and neutral, the counts of posts, sites and authors are highly correlated. Across the categories of positive, negative, and neutral, we find that positive and negative data are very highly correlated. We speculate that this may be due to the rapid responsiveness of online media. A positive or negative comment can be met immediately with an opposing viewpoint.

---Insert Tables 1 and 2 About Here---

It is useful to compare our data to that of Shin, Hanssens and Gajula (2008). For the 11 digital music players included in their study, they observed an average of 140 comments a day, 57% were positive, 5 % were negative and 38% were neutral. Considering neutral posts alone, we observe nearly 1,900 a day. Thus we see that moving from consideration of a single product to consideration of an entire firm, the volume of chatter increases dramatically. In addition, consistent with the idea that chatter for our technical company is more information-related than evaluative in nature, 79% of the total number of posts are coded as neutral while only 38% of the comments about digital music players were neutral.

In addition to the chatter data, weekly sales and gross profit margin are provided to us from the firm's internal accounting records. We use these data to briefly examine our ideas about chatter for the technical firm being correlated with either sales or gross margin. We find no significant correlation between sales and any of the chatter measures. However, we find positive and significant correlation between the total posts and gross margin ($\rho = 0.41$). Finally, weekly stock price data for our technical firm, along with industry index (NASDAQ) data are obtained from Yahoo! Finance. The weekly risk-free rate of return, measured by the weekly return on U.S. Treasury bills, is obtained from Kenneth French's data library, available at his website. Due to the timing of the 2007 New Year's holiday, we have 51 weeks of stock return data for 2007.

Models and Results

The Relationship Between Aggregate Chatter and Stock Return

Following the stock return response modeling approach of Mizik and Jacobson (2004), we assess whether the information contained in the aggregate chatter data is reflective of the information set that market participants perceive as affecting future cash flows for our firm. Consistent with Dhar and Chang (2007), we begin by considering the total counts of posts, sites and authors (e.g., aggregating across positive, negative, and neutral for each of the categories). Given the high correlation between total posts, total sites, and total authors, a naïve approach would be to discard two of the three measures and focus on only one. However, this ignores potentially useful information. To make full use of our data, we propose a confirmatory factor model to isolate the component of chatter common to the total posts, total sites, and total author counts. Consider the following system

$$
\begin{aligned}
&y_t = x_t'\beta + \xi_t\pi + \varepsilon_t \quad \varepsilon_t \sim N\left(0,\sigma_\varepsilon^2\right) \\
(1) \qquad &z_t = \mu + \Gamma z_{t-1} + \Lambda\xi_t + \omega_t \quad \omega_t \sim N\left(0,\Omega\right) \\
&\xi_t \sim N\left(0,\sigma_\xi^2\right)
\end{aligned}
$$

where $t = 1,...,T$ indexes weeks. The scalar y_t is the firm's weekly stock return less the risk-free rate of return. The vector x_t includes an intercept, the weekly market return (less the risk-free rate), and a dummy variable for the release of publicly available quarterly earnings reports. The vector z_t is a $P = 3$ dimensional column vector of total post, site, and author counts. The vector μ is a mean vector and Γ is a diagonal matrix of auto-regressive parameters. The matrix Ω is a diagonal covariance matrix. The residual covariance of the vector auto-regression, which measures the shock to chatter, is modeled via a factor structure. The vector Λ is a $P \times 1$ factor loading vector that maps the *z*-vector onto the scalar factor score ξ_t, which is the common shock to chatter. The factor score is also common to both to the auto-regression model and the stock return model. Thus, a test of whether or not chatter has value relevance is to test the hypothesis $\pi = 0$. For identification, σ_ξ^2 is set to one, which implies the covariance matrix of the residuals is given by $\Lambda\Lambda' + \Omega$. Lastly, since the likelihood is the same for $\Lambda \times \xi$ and $-(\Lambda) \times -(\xi)$, we restrict the non-zero elements of Λ to be greater than zero.

A simple approach to estimation of (1) is to first estimate the auto-regressive factor model, recover estimates of the factor scores, ξ_t, and plug these estimates into the stock return model. This two-stage estimation strategy ignores measurement error as well as the relationship between y and ξ under the model. The former may bias the estimates of π towards zero (Green 2004) while the latter may induce both bias and inefficiency (Gilbride et al. 2005). Bayesian approaches to estimation provide a natural way to deal with these problems. We can derive and sample from the full conditional distributions, including ξ, which depends on the likelihoods for both y and z.[1] We estimate equation (1) via Bayesian MCMC methods. The full conditional distributions are included in the technical appendix. For each model, the sampler is run for a total of 15,000 iterations, keeping the last 5,000 for inference.

---

[1] An alternative frequentist approach would entail choosing model parameters to maximize the multi-variate normal likelihood of *y* and *z*.

For each model, we report three measures of stock return model fit. First, we report the adjusted R^2 of the stock return regression, computed as

$1-\left(1-R^2\right)\times\left(\frac{T-1}{T-K-1}\right)$, where $R^2 = 1-\left(\frac{\sum_{t=1}^{T}\left(y_t-\hat{y}_t\right)^2}{\sum_{t=1}^{T}\left(y_t-\bar{y}\right)^2}\right)$. The scalar $\hat{y}_t$ is the posterior mean of the predicted values of y_t, where the mean is taken over iterations of the sampler. The adjusted R^2 term includes a penalty for increasing the number of regressors in the stock return model, given by K. Second, we conduct a posterior predictive check of internal validity (Gelman, Meng, and Stern 1996). Let θ^y be the parameters of the stock return model, $\left\{\beta,\xi,\pi,\sigma_\varepsilon^2\right\}$. Using the observed explanatory variables and the posterior distribution of θ^y, we generate a replicated data set, y_t^{rep}. The posterior predictive distribution of y_t^{rep} is given by $p\left(y_t^{rep}\middle|y_t\right)=\int p\left(y_t^{rep}\middle|\theta^y\right)p\left(\theta^y\middle|y_t\right)d\theta^y$

The posterior predictive distribution incorporates two kinds of uncertainty: sampling uncertainty about y given θ^y and parametric uncertainty about θ^y. Using the replicated data, we compute the root mean squared deviation (RMSD), $\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(y_t^{rep}-y_t\right)^2}$. Lastly, we report the Deviance Information Criterion (Spiegelhalter et al. 2002). The DIC is computed as $D\left(\bar{\theta}^y\right)+2p_D$, where $D\left(\bar{\theta}^y\right)=-2\log\left(f\left(y|\bar{\theta}^y\right)\right)$ is computed using the posterior means of model parameters. The effective number of model parameters is given by the term $p_D = E_{\theta|y}\left[-2\log\left(f\left(y|\theta^y\right)\right)\right]+2\log\left(f\left(y|\bar{\theta}^y\right)\right)$.

The estimation results are reported in Table 3. For comparison, we estimate a baseline model setting $\pi = 0$. To test the time series properties of the chatter auto-regression, we compute the posterior mean of the difference between 1.00 and the AR(1) parameters on the diagonal of Γ. The posterior probability that the difference is greater than zero is 1 for all three coefficients, thus we conclude that the time series behavior of chatter can be approximated by an AR(1) model. In terms of the return model, the 95% interval of highest posterior density (HPD) for the coefficient on market return (NASDAQ return less the risk-free rate) does not span zero. The posterior mean estimate of π, which measures the association between unanticipated shocks to chatter and stock return, is positive (1.046) and the 95% HPD interval, [0.078, 1.921], does not span zero. The adjusted R^2, the posterior predictive check, and the DIC of the stock return model improves with the inclusion of chatter. Altogether, the analysis yields strong evidence that unanticipated shocks to online chatter are positively associated with the firm's stock return.

---Insert Table 3 About Here---

We consider two alternative model formulations. First, we re-estimate equation system (1) including the Fama and French (1992) factors which capture market capitalization and book-to-price effects. Not surprisingly, these additional factors, designed to account for differences across firms, had no effect on our within-firm analysis. We also estimated models with a linear, a second order, and a third order polynomial trend in the stock return regression. In all cases, the 95% HPD intervals on the trend coefficients span zero. Results for these two alternative model specifications, omitted for the sake of brevity, are available from the authors upon request.

Robustness Checks

In this section, we consider several robustness checks on the main findings reported thus far. We first consider some alternative measures of abnormal stock return. We then investigate reverse causality and omitted variable bias. Lastly, we consider positive, negative and neutral chatter.

Alternative Measures of Abnormal Return

We compute two alternative measures of weekly abnormal return, the weekly cumulative abnormal return (*CAR*) and the weekly compounded abnormal return (*CPAR*). We compute CAR_t as $\sum_{m=t-5}^{t}\left[\bar{\eta}_m\right]$ and $CPAR_t$ as $\prod_{m=t-5}^{t}\left[1+\left(\bar{\eta}_m\right)\right]-1$ where $\bar{\eta}_m$ is the posterior mean of the residual term (including the constant) from a regression of daily return for the firm (less the risk-free rate) on the daily NASDAQ return (less the risk-free return). CAR_t and $CPAR_t$ are then substituted for y_t in equation (1). For this analysis, the vector x_t contains a constant and the quarterly report dummy. The results of these alternative specifications are reported in Table 4. Using these alternative measures of abnormal return, we also find a positive effect of unanticipated shocks to online chatter on stock return. Computing abnormal returns as CAR, the posterior mean estimate of π is 0.936 with the 95% HPD interval [0.138, 1.758]. Computing abnormal returns as $CPAR$, the posterior mean estimate of π is 0.971 with the 95% HPD interval [0.171, 1.834]. Again, the adjusted R^2, the posterior predictive check, and the DIC of the stock return model improves with the inclusion of the chatter variable.

---Insert Table 4 About Here---

Reverse Causality and Omitted Variable Bias

It should be noted that the stock market response modeling approach applied in this paper does not imply causation between shocks to chatter and stock return. The central idea behind our model is that the stock market cannot observe intra-quarter firm performance outcomes. It can observe chatter, which may be an indicator of current firm performance. An alternative explanation is that stock return is inducing online chatter. The relative paucity of comments that deal with financial topics casts doubt on the face validity of this explanation. Furthermore, a regression of total posts onto lagged posts and stock return yields an insignificant coefficient on stock return. Previous research has noted that in the context of the efficient markets hypothesis, instrumental variable approaches, particularly those that use lagged variables as instruments, are problematic. Rather than reverse causality, the problem may be more appropriately cast as an omitted variable problem (Mizik and Jacobson 2008).

Omitted Variables Bias-Sales and Profits

It is possible that the firm's weekly profit or sales information may reach financial markets directly, perhaps through some signaling efforts on the part of the firm. Thus, it is possible that shocks to weekly profit or sales may be driving both stock return and shocks to online chatter, and our model captures a spurious relationship due to this omitted variable. To investigate this possibility, we test whether unexpected shocks to the weekly profit or sales explain abnormal stock return by estimating the following system

(2)
$$\begin{aligned} y_t &= x_t^{'}\beta + \omega_t\gamma + \varepsilon_t \quad \varepsilon_t \sim N\left(0,\sigma_\varepsilon^2\right) \\ q_t &= \phi_0 + q_{t-1}\phi_1 + \omega_t \quad \omega_t \sim N\left(0,\sigma_\omega^2\right) \end{aligned}$$

where y_t and x_t are as before and q_t represents either the firm's weekly profit contribution or weekly sales.[2] The residual term ω_t captures the unexpected shocks to weekly profits or sales . Since $\omega_t = q_t - \phi_0 - q_{t-1}\phi_1$, the posterior distribution of the vector ϕ depends on the data likelihoods for both y_t and q_t. We also test for a relationship between stock return and unexpected shocks to profits or sales using the two alternative measures of abnormal return, cumulative abnormal return (*CAR*) and the weekly compounded abnormal return (*CPAR*).

The results of these analyses appear in Table 5a-5b. For ease of exposition, we repeat the baseline models from Tables 3 and 4 in Tables 5a and 5b. In all cases, we find the probability that the difference between 1.00 and the AR(1) parameter ϕ_1 is greater than zero to be 1. In all cases, we also find that both the 95% and 90% HPD for intervals for γ span zero. In addition, we find that including either shocks to profit or sales does not improve the fit of the stock return regression. Thus, we conclude that neither the profit nor sales information seems to be available to the market.

---Insert Tables 5a-5b About Here---

Omitted Variables Bias-New Product Launches

Similar to the effect of profit or sales data reaching the financial markets, we investigate whether information about new product launches may be driving stock return. Data on new product launch announcements are available from the firm's website. We find that new product launch announcements occurred in 23 of the weeks in our sample. Let $d_t = 1$ if a new product launch announcement occurs in week t, else $d_t = 0$. We

[2] To rule out the omitted variable bias argument, it is sufficient to show that the omitted variable is not related to stock return.

estimate the stock return equation $y_t = x_t^{'}\beta + d_t\tau + \varepsilon_t \quad \varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right)$ where y_t and x_t are as defined in equation (1). We also estimate the regression equations with weekly cumulative abnormal return (*CAR*) and the weekly compounded abnormal return (*CPAR*) as the dependent variable. The results appear in Table 6. In all cases, we find that both the 90% and 95% HPD for intervals for τ span zero. In addition, we find that including the new product launch announcement indicator variable does not improve the fit of any of the models.[3]

---Insert Table 6 About Here---

A Deeper Examination of Positive and Negative Chatter

In contrast to Dhar and Chang (2007), some previous research has documented that positive and negative word of mouth can have differential effects on firm outcomes (Shin, Hanssens and Gajula 2008, Luo 2007, 2008). As discussed and shown in Table 2, the positive and negative dimensions of our chatter measures are highly correlated. We examine whether reducing the dimensionality of the data via factor analysis might result in a positive and negative factor scores that exhibit less correlation. To accommodate this effort, we expand the z_t vector in equation (1) to include the counts of posts, sites, and authors for positive, negative and neutral dimensions, such that $P = 9$. The model is specified as

$$(3) \qquad \begin{aligned} & y_t = x_t^{'}\beta + \xi_t^{'}\pi + \varepsilon_t \quad \varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right) \\ & z_t = \mu + \Gamma z_{t-1} + \Lambda \xi_t + \omega_t \quad \omega_t \sim N\left(0, \Omega\right) \\ & \xi_t \sim N\left(0, \Xi\right) \end{aligned}$$

[3] We also experimented with windows of differing sizes around each of the launch announcements. Some of the announcements were in adjacent weeks, thus the windows for one announcement often overlap with subsequent announcements. Using windows instead of a simple dummy, we find no significant effect of announcements in any of the models.

where Ω is a diagonal covariance matrix. The model for z_t is a confirmatory factor model, with the patterned $P \times M$ matrix Λ mapping the vector z_t onto the $M = 3$ dimensional vector ξ_t , which measures signed chatter, positive, negative and neutral. The correlation matrix Ξ measures the correlation in the latent factors. Estimation proceeds as before, with the only complication arising from estimation of the correlation matrix, Ξ , as no convenient prior exists for correlation matrices. To navigate around this issue, we sample from the unidentified covariance matrix and margin down to the identified parameters (Edwards and Allenby 2004).

We estimate equation (3) with weekly returns as the dependent variable in the stock return regression, as well as with weekly cumulative and compounded abnormal returns as the dependent variable. The results are reported in Table 7. For all three models, the signs on positive and negative chatter parameters are positive and negative, respectively, but both the 95% and the 90% HPD intervals span zero. For the model with weekly returns as the dependent variable, the posterior mean estimate of the parameter on neutral chatter, π^{NEU} , is 0.919; the 90% HPD interval, [0.087, 1.785], does not span zero. For the cumulative abnormal returns model, the posterior mean estimate of π^{NEU} is 0.990; the 95% HPD interval, [0.030, 1.962], does not span zero. For the compounded abnormal returns model, the posterior mean estimate of π^{NEU} is 0.993; the 95% HPD interval, [0.039, 1.998], does not span zero. For all three models, the fit statistics are improved by the inclusion of the chatter shocks.

---Insert Table 7 About Here---

Table 8 reports the posterior mean estimate of the correlation matrix, Ξ . The correlation between ξ_t^{POS} and ξ_t^{NEG} is 0.81. The high correlation between ξ_t^{POS} and ξ_t^{NEG}

suggests that the classic problem of multi-collinearity may be affecting our ability to estimate the parameters on these variables in the stock return regression. To investigate this further, we now turn our attention to a higher order factor representation of the online chatter data.

---Insert Table 8 About Here---

We model the correlation between the positive and negative factor scores via a higher order factor model (Yung et al. 1999). The general form of the model is as follows

$$\begin{aligned} &y_t = x_t^{'}\beta + \upsilon_t^{*'}\pi + \varepsilon_t \quad \varepsilon_t \sim N\left(0,\sigma_\varepsilon^2\right) \\ &z_t = \mu + \Gamma z_{t-1} + \Lambda\xi_t + \omega_t \quad \omega_t \sim N\left(0,\Omega_\omega\right) \\ &\xi_t = \Psi\eta_t + \upsilon_t \quad \upsilon_t \sim N\left(0,\Omega_\upsilon\right) \\ &\eta_t \sim N\left(0,\sigma_\eta^2\right) \end{aligned} \qquad (4)$$

Here, the correlated positive and negative chatter factor scores are modeled as arising from a higher order, scalar factor score common to positive and negative chatter, η_t and a residual term υ_t with diagonal covariance matrix Ω_υ. The vector $\Psi = \left[\psi_{POS}\ \psi_{NEG}\ 0\right]^{'}$ maps the η_t onto ξ_t. Note the zero restriction, which restricts the effect of η to the positive and negative dimensions of chatter. For identification, the variance σ_η^2 is set to 1. This implies the correlation matrix of ξ_t is given by $\Psi\Psi^{'} + \Omega_\upsilon$. Given the restriction that the diagonal elements of $\Psi\Psi^{'} + \Omega_\upsilon$ equal 1, Ω_υ is computed directly from the restriction $I_m = diag\left(\Psi\Psi^{'}\right) + \Omega_\upsilon$. Since the diagonal elements of Ω_υ must be greater than zero, this implies the restriction $1-\psi_m^2 > 0$ where ψ_m is the m^{th} element of Ψ.

The stock return is modeled as a function of the normalized residuals υ_t^*, where $\upsilon_{tm}^* = \frac{\upsilon_{tm}}{\sqrt{\Omega_{\upsilon m}}}$. This normalization ensures that each element of υ_t^* has unit variance. In this specification, we have isolated the orthogonal components of positive and negative chatter. The question is whether or not any signal remains in these components. We estimate equation (4) with weekly returns as the dependent variable in the stock return regression, as well as with weekly cumulative and compounded abnormal returns as the dependent variable. The results are presented in Table 9. Both the 95% and 90% HPD intervals span zero for π^{POS}, and π^{NEG} for all three stock return regressions. However, the standard errors of the estimates are smaller compared to those reported in Table 8. For each of the three regressions, π^{NEU} is positive and the 95% HPD intervals do not span zero. Cleaning out the common component of positive and negative chatter does seem to reduce the noise in the data, thereby increasing model fit statistics. However, only neutral chatter has an effect on stock return.

---Insert Table 9 About Here---

Summary and Conclusions

Previous research has shown that unexpected shocks in firm descriptors available to the market via formal reporting mechanisms (e.g., publicly available quarterly reports) can impact stock return. Mizik and Jacobson (2004) discuss other channels by which the market receives information about firms. Our analysis considers one such channel, the visible trace of online communications about a firm and its products and services. The marketing literature has demonstrated a link between such online communications, or chatter, and sales and prices. Given this established link between chatter and firm performance, the observability of chatter, and the availability of chatter on an intra-quarter level, it is interesting to consider whether or not measures of chatter provide added explanatory power to standard variables used to explain stock price movements. If so, we can conclude that the market views online communications as indicative of information perceived to affect future cash flows, or value relevant (Mizik and Jacobson 2004).

Our measures of chatter are provided to us by a technical products firm. The measures consider all mentions of the firm, its products and services. At the firm level, we find that shocks to total weekly chatter are associated with stock returns. In addition, we find that this result is robust. The relationship between chatter and firm value holds for alternative specifications of the model (e.g., the Fama-French three factor model and a model that includes linear and nonlinear time effects) and for alternative measures of stock return (e.g., cumulative abnormal returns and compounded abnormal returns). The data do not support a reverse causation argument nor do they support sales, profit or new product launches as omitted variables that might be driving both chatter and stock return.

Considering the positive, negative, and neutral components of chatter, a challenge arises from the high correlation between positive and negative chatter. We argue that such high correlation is a symptom of the nature of online communication, where praise can be met swiftly with scorn, and vice-versa. We propose a higher order factor model to isolate the orthogonal components of positive and negative chatter and still find no significant effect of positive and negative chatter. Only shocks to neutral chatter that are associated with stock returns.

With this paper we add to the literature linking online chatter and market response. Earlier literature linked chatter about specific products to the product's sales or price performance. We extend these findings by considering the total amount of chatter about the firm and its potential relationship with stock returns. While the paper's findings are intriguing, future research should replicate, extend and find boundary conditions for these results. Should one expect these results to hold for a non-technical firm whose transactions may not trigger a flurry of online interactions? What is the sensitivity of these findings to the larger economic environment? Should the effect continue to hold during the market disruption we are experiencing in 2008-2009? Finally, the Internet continues to evolve. New communication vehicles arise and some grow explosively (e.g., Twitter). Financial markets are notoriously efficient in their consumption of information about firm performance. Surely the use to which financial markets put this information will evolve as the information generating process itself evolves.

References

Antweiler, W., M. Frank (2004), "Is all that talk just noise? The information content of Internet stock message boards", *Journal of Finance* 59, 1259-1293.

Chevalier, Judith A. and Dina Mayzlin (2006), "The Effect of Work of Mouth on Sales: Online Book Reviews", *Journal of Marketing Research* 43 (August), 345-354.

Cutler, David M., James M. Poterba, and Lawrence H. Summers (1989), "What moves stock prices?", *Journal of Portfolio Management* 15, 4-12.

Das, S. and M. Chen (2004), "Yahoo! For Amazon: Sentiment Extraction from Small Talk on the Web," Santa Clara University Working Paper.

Dellarocas, C. , M. Zhang and N.F. Awad (2007), "Exploring the Value of Online Product Reviews in Forecasting Sales: The Case of Motion Pictures", *Journal of Interactive Marketing* 21 (4, Fall), 23-45.

Dhar, Vasant and Elaine Chang (2007), "Does Chatter Matter? The Impact of User-Generated Content on Music Sales", Working Paper Leonard N. Stern School of Business, New York University.

Edwards, Y. and G. Allenby (2003), "Multivariate Analysis of Multiple Response Data", *Journal of Marketing Research*, 40, 321-334.

Fama, E. and K. French (1992), "The Cross-Section of Expected Stock Returns", *Journal of Finance* **47** (2): 427–465

Gelman, A., X. L. Meng, and H. S. Stern (1996), "Posterior Predictive Assessment of Model Fitness via Realized Discrepancies", *Statistica Sinica*, Vol. 6, No. 4, 733-807.

Godes, David and Dina Mayzlin (2004), "Using Online Conversations to Study Word-of-Mouth Communication", *Marketing Science* 23 (4, Fall), 545-560.

Godes, David, Dina Mayzlin, Yubo Chen, Sanjiv Das, Chrysanthos Dellarocas, Bruce Pfeiffer, Barak Libai, Subrata Sen, Meng Shi, Peeter Verlegh (2005), "The Firm's Management of Social Interactions", *Marketing Letters* 16 (3/4), 415-428.

Gilbride T., S. Yang and G. Allenby (2005), "Modeling Simultaneity in Survey Data", *Quantitative Marketing and Economics*, 3, 311-335.

Green, W., (2000), *Econometric Analysis*, New Jersey: Prentice-Hall.

Liu, Yong (2006), "Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue", *Journal of Marketing* 70 (July), 74-89.

Luo, Xueming (2007), "Consumer Negative Voice and Firm-Idiosyncratic Stock Returns", *Journal of Marketing* 71 (July), 75-88.

---(2008), "Quantifying the Long-Term Impact of Negative Word of Mouth on Cash Flows and Stock Prices", *Marketing Science* (forthcoming).

Mizik, N. and R. Jacobson (2004), "Stock Return Response Modeling", in *Assessing Marketing Strategy Performance*, C. Moorman and D. Lehmann, eds. Boston: Marketing Science Institute, 29-46.

---and R. Jacobson (2008), "The Financial Value Impact of Perceptual Brand Attributes", *Journal of Marketing Research*, 45 (February), 15-32.

Rist, P. (2005), "BIGresearch Releases SIMM VII; Word of Mouth Most Influential, Other Media Vary by Demos and Product Categories", ***www.bigresearch.com***, accessed on July 14, 2009, <http://www.bigresearch.com/news/big122005.htm>.

Senecal, S. and Nantel, J (2004), "The Influence of Online Product Recommendations on Consumers' Online Choices", *Journal of Retailing* 80, 159-69.

Shin, Hyun S., Dominique M. Hanssens and Bharath Gajula (2008), "The Impact of Positive vs. Negative Online Buzz on Retail Prices". UCLA Working Paper.

Spiegelhalter, David J., Nicola G. Best, Bradley P. Carlin, and Angelika van der Linde (2002), "Bayesian Measure of Model Complexity and Fit", *Journal of Royal Statistics Society*, Vol. 64, No. 4, 583-639.

Tetlock, Paul C, (2007), "Giving Content to Investor Sentiment: The Role of Media in the Stock Market". *Journal of Finance* 42 (3, June), 1139-1168.

---, M. Saar-Tsechansky, S. Macskassy (2007), "More than words: Quantifying language to measure firms' fundamentals. Working paper University of Texas, Austin, TX.

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

Wysocki , P (1999), "Cheap talk on the web: The determinants of postings on stock message boards." Working Paper 98025, University of Michigan Business School, Ann Arbor, MI.

Yung, Y., D. Thissen, and L. McLeod, (1999) "On the Relationship Between the Higher-Order Factor Model and the Hierarchical Factor Model", Psychometrika, 64, 112-128.

Table 1. Descriptive Statistics, Chatter Data[a]

		Mean	Median	Standard Deviation
POS[b]	Pst[c]	1.44	1.11	1.30
	Sit	1.04	0.80	0.91
	Aut	1.16	0.95	0.99
NEG	Pst	2.13	1.60	1.83
	Sit	1.33	1.12	0.97
	Aut	1.65	1.31	1.29
NEU	Pst	13.22	11.13	9.07
	Sit	7.85	8.23	4.88
	Aut	10.14	9.25	6.62

a. All data in 1,000's
b. POS indicates positive, NEG indicates negative, and NEU indicates neutral
c. Pst indicates posts, Sit indicates sites, and Aut indicates authors

Table 2. Correlation Matrix, Chatter Data

		POS[a]			NEG			NEU		
		Pst[b]	Sit	Aut	Pst	Sit	Aut	Pst	Sit	Aut
POS	Pst	1.00								
	Sit	0.95*	1.00							
	Aut	0.95*	0.99*	1.00						
NEG	Pst	0.79*	0.79*	0.80*	1.00					
	Sit	0.90*	0.94*	0.95*	0.91*	1.00				
	Aut	0.85*	0.88*	0.88*	0.96*	0.97*	1.00			
NEU	Pst	0.20	0.09	0.12	0.29	0.19	0.25	1.00		
	Sit	0.36	0.29	0.31	0.41	0.37	0.40	0.90	1.00	
	Aut	0.27	0.20	0.22	0.36	0.29	0.34	0.95*	0.99*	1.00

a. POS indicates positive, NEG indicates negative, and NEU indicates neutral
b. Pst indicates posts, Sit indicates sites, and Aut indicates authors
* indicates correlation is significant at the 0.05 level

Table 3. Stock Return Regressions, With and Without Chatter Shocks[a,b]

	Return Regression without Chatter	*Return Regression With Chatter*
CONSTANT	-0.543	-0.115
	(1.640)	(1.641)
NASDAQ	0.962***	1.011***
	(0.167)	(0.167)
QTR REPORT	-0.882	-0.497
	(1.623)	(1.552)
Δ *CHATTER*[c]	---	0.971**
		(0.465)
σ_{ε}^{2}	9.271	8.679
	(2.011)	(1.865)
Adjusted R^2	0.40	0.46
RMSD	2.76	2.60
DIC	251.88	248.44

a. Dependent variable is weekly stock return less the risk-free rate. The return is computed as $\frac{pri_t - pri_{t-1}}{pri_{t-1}}$ where pri_t is the stock price at week *t*.

b. Cell entries are posterior means and posterior standard errors (in parentheses)

c. Δ *CHATTER* is the factor score from the auto-regression in equation (1)

*** indicates the 99% highest posterior density interval does not span zero

** indicates the 95% highest posterior density interval does not span zero

* indicates the 90% highest posterior density interval does not span zero

Table 4. Alternative Abnormal Return Regressions, With and Without Chatter Shocks[a,b]

	CAR Regression		*CPAR Regression*	
	Without Chatter	*With Chatter*	*Without Chatter*	*With Chatter*
CONSTANT	0.682	0.652	0.683	0.652
	(0.414)	(0.416)	(0.427)	(0.416)
QTR REPORT	-1.075	-0.606	-1.028	-0.606
	(1.477)	(1.423)	(1.507)	(1.423)
Δ *CHATTER*[c]	---	0.936**	---	0.936**
		(0.410)		(0.410)
σ_ε^2	8.050	7.731	8.381	7.467
	(1.718)	(1.626)	(1.708)	(1.573)
Adjusted R^2	-0.03	0.07	-0.03	0.08
RMSD	2.67	2.51	2.72	2.56
DIC	246.25	242.30	247.99	244.12

a. CAR_t as $\sum_{m=t-5}^{t}\left[\bar{\eta}_m\right]$ and $CPAR_t$ as $\prod_{m=t-5}^{t}\left[1+\left(\bar{\eta}_m\right)\right]-1$ where $\bar{\eta}_m$ is the posterior mean of the residual term (including the constant) from a regression of daily return for the firm (less the risk-free rate) on the daily NASDAQ return (less the risk-free return).

b. Cell entries are posterior means and posterior standard errors (in parentheses)

c. Δ *CHATTER* is the factor score from the auto-regression in equation (1)

*** indicates the 99% highest posterior density interval does not span zero

** indicates the 95% highest posterior density interval does not span zero

* indicates the 90% highest posterior density interval does not span zero

Table 5a. Stock Return Regressions, With and Without Profit Shocks[a,b]

	Dependent Variable					
	RET		*CAR*		*CPAR*	
	Without Profits	*With Profits*	*Without Profits*	*With Profits*	*Without Profits*	*With Profits*
CONSTANT	-0.543	-0.492	0.682	0.680	0.683	0.682
	(1.640)	(1.645)	(0.414)	(0.420)	(0.427)	(0.437)
NASDAQ	0.962***	0.967***	---	---	---	---
	(0.167)	(0.168)				
QTR REPORT	-0.882	-0.893	-1.075	-1.019	-1.028	-1.001
	(1.623)	(1.585)	(1.477)	(1.466)	(1.507)	(1.498)
Δ *PROFITS*[c]	---	0.016	---	0.029	---	0.029
		(0.026)		(0.024)		(0.024)
σ_{ε}^{2}	9.271	9.028	8.050	7.800	8.381	8.140
	(2.011)	(1.922)	(1.718)	(1.632)	(1.708)	(1.726)
Adjusted R^2	0.40	0.40	-0.03	-0.01	-0.03	-0.01
RMSD	2.76	2.74	2.67	2.61	2.72	2.67
DIC	251.88	253.01	246.25	246.13	247.99	248.11

a. *RET* is weekly stock return less the risk-free rate. The return is computed as $\frac{pri_t - pri_{t-1}}{pri_{t-1}}$ where pri_t is the stock price at week *t*. CAR_t as $\sum_{m=t-5}^{t}\left[\bar{\eta}_m\right]$ and $CPAR_t$ as $\prod_{m=t-5}^{t}\left[1+\left(\bar{\eta}_m\right)\right]-1$ where $\bar{\eta}_m$ is the posterior mean of the residual term (including the constant) from a regression of daily return for the firm (less the risk-free rate) on the daily NASDAQ return (less the risk-free return)

b. Cell entries are posterior means and posterior standard errors (in parentheses)

c. Δ *PROFITS* is the residual term from the auto-regression in equation (3).

*** indicates the 99% highest posterior density interval does not span zero

** indicates the 95% highest posterior density interval does not span zero

* indicates the 90% highest posterior density interval does not span zero

Table 5b. Stock Return Regressions, With and Without Sales Shocks[a,b]

	Dependent Variable					
	RET		*CAR*		*CPAR*	
	Without Sales	*With Sales*	*Without Sales*	*With Sales*	*Without Sales*	*With Sales*
CONSTANT	-0.543	-0.736	0.682	0.688	0.683	0.677
	(1.640)	(1.657)	(0.414)	(0.428)	(0.427)	(0.430)
NASDAQ	0.962***	0.944***	---	---	---	---
	(0.167)	(0.170)				
QTR REPORT	-0.882	-0.648	-1.075	-0.962	-1.028	-0.984
	(1.623)	(1.582)	(1.477)	(1.475)	(1.507)	(1.524)
$\Delta SALES$[c]	---	-0.563	---	-0.266	---	-0.274
		(0.644)		(0.606)		(0.631)
σ_ε^2	9.271	8.952	8.050	8.123	8.381	8.475
	(2.011)	(1.897)	(1.718)	(1.700)	(1.708)	(1.850)
Adjusted R^2	0.40	0.40	-0.03	-0.05	-0.03	-0.05
RMSD	2.76	2.73	2.67	2.66	2.72	2.71
DIC	251.88	252.74	246.25	247.90	247.99	249.90

a. *RET* is weekly stock return less the risk-free rate. The return is computed as $\frac{pri_t - pri_{t-1}}{pri_{t-1}}$ where pri_t is the stock price at week *t*. CAR_t as $\sum_{m=t-5}^{t}\left[\bar{\eta}_m\right]$ and $CPAR_t$ as $\prod_{m=t-5}^{t}\left[1+\left(\bar{\eta}_m\right)\right]-1$ where $\bar{\eta}_m$ is the posterior mean of the residual term (including the constant) from a regression of daily return for the firm (less the risk-free rate) on the daily NASDAQ return (less the risk-free return)

b. Cell entries are posterior means and posterior standard errors (in parentheses)

c. $\Delta SALES$ is the residual term from the auto-regression in equation (3).

*** indicates the 99% highest posterior density interval does not span zero

** indicates the 95% highest posterior density interval does not span zero

* indicates the 90% highest posterior density interval does not span zero

Table 6. Stock Return Regressions, With and Without New Product Launch Announcements[a,b]

	Dependent Variable					
	RET		*CAR*		*CPAR*	
	Without NPL	*With NPL*	*Without NPL*	*With NPL*	*Without NPL*	*With NPL*
CONSTANT	-0.543	-0.282	0.682	0.811	0.683	0.808
	(1.640)	(1.910)	(0.414)	(0.571)	(0.427)	(0.587)
NASDAQ	0.962***	0.978	---	---	---	---
	(0.167)	(0.180)				
QTR REPORT	-0.882	-0.912	-1.075	-1.021	-1.028	-1.010
	(1.623)	(1.650)	(1.477)	(1.530)	(1.507)	(1.520)
NEW PRODUCT LAUNCH (NPL) ANNOUNCEMENT[c]	---	-0.274	---	-0.263	---	-0.273
		(0.912)		(0.828)		(0.857)
σ_{ε}^{2}	9.271	9.667	8.050	8.460	8.381	8.763
	(2.011)	(2.064)	(1.718)	(1.798)	(1.708)	(1.858)
Adjusted R^2	0.40	0.39	-0.03	-0.05	-0.03	-0.05
RMSD	2.76	2.76	2.67	2.67	2.72	2.71
DIC	251.88	253.97	246.25	248.34	247.99	250.17

a. *RET* is weekly stock return less the risk-free rate. The return is computed as $\frac{pri_t - pri_{t-1}}{pri_{t-1}}$ where pri_t is the stock price at week *t*. CAR_t as $\sum_{m=t-5}^{t}\left[\bar{\eta}_m\right]$ and $CPAR_t$ as $\prod_{m=t-5}^{t}\left[1+\left(\bar{\eta}_m\right)\right]-1$ where $\bar{\eta}_m$ is the posterior mean of the residual term (including the constant) from a regression of daily return for the firm (less the risk-free rate) on the daily NASDAQ return (less the risk-free return)

b. Cell entries are posterior means and posterior standard errors (in parentheses)

c. *NEW PRODUCT LAUNCH (NPL) ANNOUNCEMENT* is dummy indicator

*** indicates the 99% highest posterior density interval does not span zero

** indicates the 95% highest posterior density interval does not span zero

* indicates the 90% highest posterior density interval does not span zero

Table 7. Stock Return Regressions, With and Without Signed Chatter Shocks[a,b]

	Dependent Variable					
	RET		*CAR*		*CPAR*	
	Without Chatter	*With Chatter*	*Without Chatter*	*With Chatter*	*Without Chatter*	*With Chatter*
CONSTANT	-0.543	0.036	0.682	0.677	0.683	0.636
	(1.640)	(1.690)	(0.414)	(0.454)	(0.427)	(0.453)
NASDAQ	0.962***	1.027***	---	---	---	---
	(0.167)	(0.173)				
QTR REPORT	-0.882	-0.578	-1.075	-0.685	-1.028	-0.648
	(1.623)	(1.603)	(1.477)	(1.480)	(1.507)	(1.487)
SIGNED CHATTER SHOCKS						
Δ *Positive*[c]	---	0.389	---	0.552	---	0.633
		(1.381)		(1.070)		(1.153)
Δ *Negative*	---	-0.215	---	-0.462	---	-0.524
		(1.359)		(1.065)		(1.143)
Δ *Neutral*	---	0.919*	---	0.990**	---	0.993**
		(0.520)		(0.491)		(0.498)
σ_{ε}^{2}	9.271	8.561	8.050	7.531	8.381	7.694
	(2.011)	(2.015)	(1.718)	(1.703)	(1.708)	(1.782)
Adjusted R^2	0.40	0.49	-0.03	0.12	-0.03	0.14
RMSD	2.76	2.48	2.67	2.39	2.72	2.41
DIC	251.88	248.74	246.25	243.12	247.99	243.84

a. *RET* is weekly stock return less the risk-free rate. The return is computed as $\frac{pri_t - pri_{t-1}}{pri_{t-1}}$ where pri_t is the stock price at week *t*. CAR_t as $\sum_{m=t-5}^{t}\left[\bar{\eta}_m\right]$ and $CPAR_t$ as $\prod_{m=t-5}^{t}\left[1+\left(\bar{\eta}_m\right)\right]-1$ where $\bar{\eta}_m$ is the posterior mean of the residual term (including the constant) from a regression of daily return for the firm (less the risk-free rate) on the daily NASDAQ return (less the risk-free return)

b. Cell entries are posterior means and posterior standard errors (in parentheses)

c. Δ *Positive*, Δ *Negative*, and Δ *Neutral* are the factor scores from equation (4).

*** indicates the 99% highest posterior density interval does not span zero
** indicates the 95% highest posterior density interval does not span zero
* indicates the 90% highest posterior density interval does not span zero

Table 8. Factor Score Correlation Matrix[a]

	ξ_t^{POS}	ξ_t^{NEG}	ξ_t^{NEU}
ξ_t^{POS}	1.00	0.81***	0.26*
ξ_t^{NEG}	0.81***	1.00	0.33*
ξ_t^{NEU}	0.26*	0.33*	1.00

a. Estimated with weekly returns as the dependent variable in the stock return regression.
*** indicates the 99% highest posterior density interval does not span zero
** indicates the 95% highest posterior density interval does not span zero
* indicates the 90% highest posterior density interval does not span zero

Table 9. Stock Return Regressions, With and Without Signed Chatter Shocks[a,b]

	Dependent Variable					
	RET		*CAR*		*CPAR*	
	Without Chatter	*With Chatter*	*Without Chatter*	*With Chatter*	*Without Chatter*	*With Chatter*
CONSTANT	-0.543	0.118	0.682	0.595	0.683	0.586
	(1.640)	(1.614)	(0.414)	(0.453)	(0.427)	(0.457)
NASDAQ	0.962***	1.035***	---		---	
	(0.167)	(0.164)				
QTR REPORT	-0.882	-0.585	-1.075	-0.659	-1.028	-0.597
	(1.623)	(1.586)	(1.477)	(1.482)	(1.507)	(1.478)
SIGNED CHATTER SHOCKS						
Δ *Positive*[c]	---	0.235	---	0.361	---	0.436
		(0.971)		(0.918)		(0.964)
Δ *Negative*	---	-0.001	---	-0.207	---	-0.053
		(0.942)		(0.901)		(0.925)
Δ *Neutral*	---	0.997**	---	0.964**	---	0.998**
		(0.486)		(0.440)		(0.448)
σ_{ε}^{2}	9.271	7.239	8.050	6.290	8.381	6.391
	(2.011)	(2.195)	(1.718)	(2.004)	(1.708)	(2.158)
Adjusted R^2	0.40	0.63	-0.03	0.39	-0.03	0.41
RMSD	2.76	2.11	2.67	2.00	2.72	2.00
DIC	251.88	230.99	246.25	223.55	247.99	222.88

a . *RET* is weekly stock return less the risk-free rate. The return is computed as $\frac{pri_t - pri_{t-1}}{pri_{t-1}}$ where pri_t is the stock price at week *t*. CAR_t as $\sum_{m=t-5}^{t}\left[\bar{\eta}_m\right]$ and $CPAR_t$ as $\prod_{m=t-5}^{t}\left[1+\left(\bar{\eta}_m\right)\right]-1$ where $\bar{\eta}_m$ is the posterior mean of the residual term (including the constant) from a regression of daily return for the firm (less the risk-free rate) on the daily NASDAQ return (less the risk-free return)

b. Cell entries are posterior means and posterior standard errors (in parentheses)

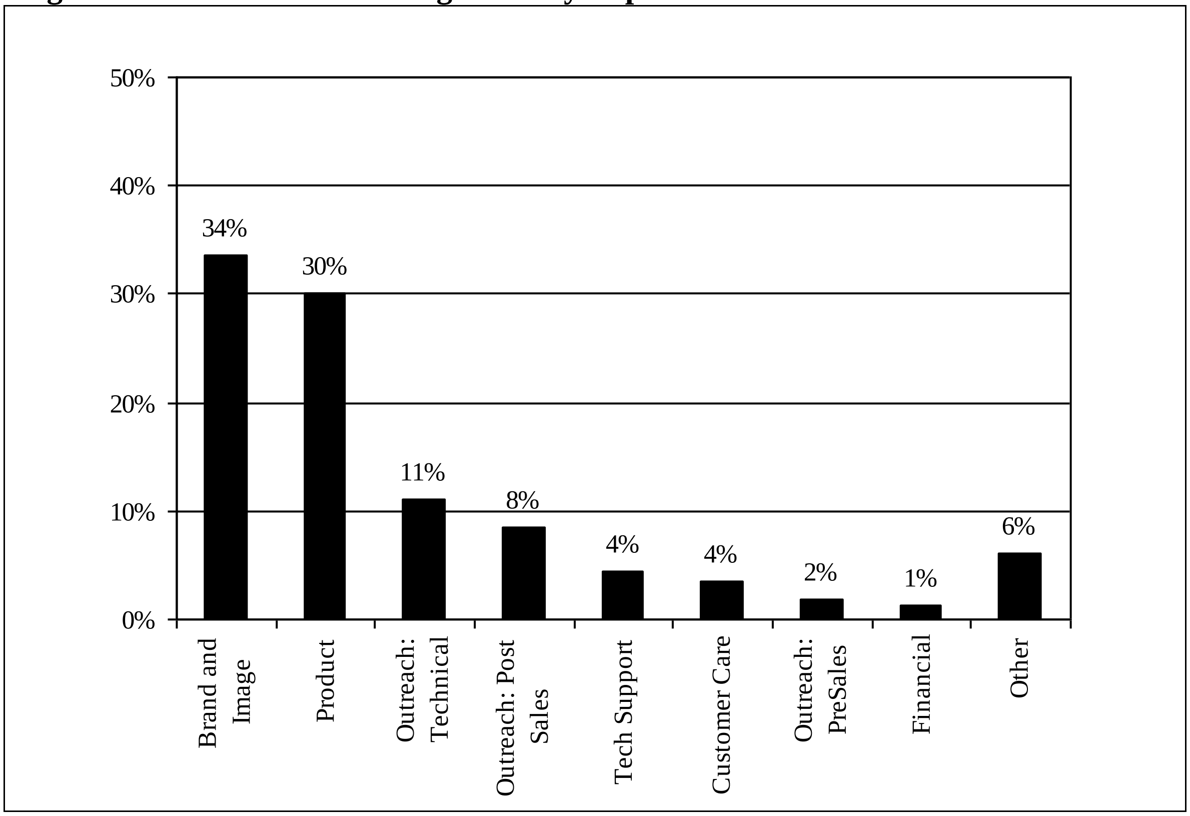
c. The Δ *Positive*, *Negative* and *Neutral Chatter* variables are the normalized residual terms from the higher order factor model in (5).

*** indicates the 99% coverage interval does not span zero

** indicates the 95% coverage interval does not span zero

* indicates the 90% coverage interval does not span zero

Figure 1. Online Chatter Categorized by Topic



Technical Appendix

This appendix details the MCMC sampler for the model in equation (1). Let $\theta = (\beta', \pi)'$ and $\xi = (\xi_1, \ldots, \xi_T)'$. The priors for $\theta, \sigma_\varepsilon^2$, μ, Γ, Λ, ξ, and Ω are given below. Each prior is proper but noninformative.

(1) $\theta \sim N(0, cI)$ with $c = 10^6$;

(2) $\sigma_\varepsilon^2 \sim IG(a_\sigma, b_\sigma)$ with $a_\sigma = 2 + K$ and $b_\sigma = 1$ where K is the number of elements in θ;

(3) $\mu \sim N(0, cI)$ with $c = 10^6$;

(4) $\gamma_p \sim N(0, c)$ where γ_p is the p^{th} diagonal element of Γ and $c = 10^6$;

(5) $\lambda_p \sim Lognormal(0, c)$ where λ_p is the p^{th} diagonal element of Λ and $c = 10^6$;

(6) $\xi_t \sim N(0, 1)$;

(7) $\omega_p \sim IG(a_\omega, b_\omega)$ where ω_p is the p^{th} diagonal element of Ω, and $a_\omega = 2$ and $b_\omega = 1$.

The MCMC algorithm requires generating random variates from the following conditional distributions. Let $y = (y_1, \ldots, y_T)'$, $X = (x_1', \ldots, x_T')'$ and $Z = (z_1', \ldots, z_T')'$.

1. The full conditional distribution for $\theta \mid y, Z, \sigma_\varepsilon^2, \mu, \Gamma, \Lambda, \xi, \Omega$ is

$$N\left[\left(\frac{1}{c}I + \frac{1}{\sigma_\varepsilon^2}\tilde{X}'\tilde{X}\right)^{-1}\left(\frac{1}{\sigma_\varepsilon^2}\tilde{X}'y\right), \left(\frac{1}{c}I + \frac{1}{\sigma_\varepsilon^2}\tilde{X}'\tilde{X}\right)^{-1}\right]$$

   where $\tilde{X} = (X, \xi)$.

2. The full conditional distribution for $\sigma_\varepsilon^2 \mid y, Z, \theta, \mu, \Gamma, \Lambda, \xi, \Omega$ is

$$IG\left(a_\sigma + \frac{1}{2}(y - \tilde{X}\theta)'(y - \tilde{X}\theta),\ b_\sigma + \frac{T}{2}\right)$$

3. The full conditional distribution for $\mu \mid y, Z, \theta, \sigma_\varepsilon^2, \Gamma, \Lambda, \xi, \Omega$ is

$$N\left[\left(\frac{1}{c} + \frac{T}{\omega_p}\right)^{-1}\left(\frac{1}{\omega_p}\sum_{t=2}^{T}\tilde{z}_{pt}\right), \left(\frac{1}{c} + \frac{T}{\omega_p}\right)^{-1}\right]$$

where z_{pt} is the p^{th} element of z_t, and $\tilde{z}_{pt} = z_{pt} - (\gamma_p z_{p,t-1} + \lambda_p \xi_t)$ and $z_{p0} = 0$.

4. The full conditional distribution for $\gamma_p \mid y, Z, \theta, \sigma_\varepsilon^2, \mu, \Lambda, \xi, \Omega$ is

$$N\left[\left(\frac{1}{c}+\frac{1}{\omega_p}\sum_{t=2}^{T} z_{p,t-1}^2\right)^{-1}\left(\frac{1}{\omega_p}\sum_{t=2}^{T} z_{p,t-1}\tilde{z}_{pt}\right),\left(\frac{1}{c}+\frac{1}{\omega_p}\sum_{t=2}^{T} z_{p,t-1}^2\right)^{-1}\right]$$

where $\tilde{z}_{pt} = z_{pt} - (\mu_p + \lambda_p \xi_t)$ and μ_p is the p^{th} element of μ.

5. The full conditional distribution for $\lambda_p \mid y, Z, \theta, \sigma_\varepsilon^2, \mu, \Gamma, \xi, \Omega$ is proportional to

$$\frac{1}{\lambda_p}\exp\left[\frac{-1}{2\omega_p^2}\sum_{t=2}^{T}(\tilde{z}_{pt}-\lambda_p\xi_t)^2\right]\exp\left\{\frac{-1}{2c}[\log(\lambda_p)]^2\right\}$$

where $\tilde{z}_{pt} = z_{pt} - (\mu_p + \gamma_p z_{p,t-1})$. This is a non-standard distribution so a random walk Metropolis-Hastings algorithm is used to generate observations.

6. The full conditional distribution for $\xi_t \mid y, Z, \theta, \sigma_\varepsilon^2, \mu, \Gamma, \Lambda, \Omega$ is

$$N\left[\left(1+\frac{1}{\sigma_\varepsilon^2}\pi^2+\Lambda'\Omega^{-1}\Lambda\right)^{-1}\left(\frac{1}{\sigma_\varepsilon^2}\pi\tilde{y}_t+\Lambda'\Omega^{-1}\tilde{z}_t'\right),\left(1+\frac{1}{\sigma_\varepsilon^2}\pi^2+\Lambda'\Omega^{-1}\Lambda\right)^{-1}\right]$$

where $\tilde{y}_t = y_t - x_t'\beta$ and $\tilde{z}_{pt} = z_{pt} - (\mu_p + \gamma_p z_{p,t-1})$.

7. The full conditional distribution for $\omega_p \mid y, Z, \theta, \sigma_\varepsilon^2, \mu, \Gamma, \Lambda, \xi$ is

$$IG\left(a_\omega+\frac{1}{2}\sum_{t=2}^{T}\tilde{z}_{pt}^2,\ b_\omega+\frac{T}{2}\right)$$

where $\tilde{z}_{pt} = z_{pt} - (\mu_p + \gamma_p z_{p,t-1} + \lambda_p \xi_t)$