

**Listening in on Online Conversations:
Measuring Brand Sentiment with Social Media**

David A. Schweidel
Assistant Professor of Marketing
Wisconsin School of Business
University of Wisconsin – Madison
975 University Avenue
Madison, WI 53706
(608) 265-2736
dschweidel@bus.wisc.edu

Wendy W. Moe
Associate Professor of Marketing
Robert H. Smith School of Business
University of Maryland
3469 Van Munching Hall
College Park, MD 20472
(301) 405-9187
wmoe@rhsmith.umd.edu

Chris Boudreaux
Sr. Vice President of Business Integration
Converseon, Inc.
53 West 38th Street, 8th Floor
New York, NY 10018
cboudreaux@converseon.com

June 2011

**Listening in on Online Conversations:
Measuring Brand Sentiment with Social Media**

ABSTRACT

With the proliferation of social media, questions have begun to emerge about its role in firms' marketing research programs. In this research, we investigate the potential to "listen in" on social media conversations as a means of inferring brand sentiment. Our analysis employs data collected from multiple website domains, spanning a variety of online venue formats to which social media comments may be contributed. We demonstrate how factors relating to the focus of social media comments and the venue to which they have been contributed can be explicitly modeled in an effort to derive a measure of online brand sentiment. Our proposed approach provides an adjusted brand sentiment metric that is highly correlated with the results of an offline brand tracking survey. This is in stark contrast to the virtually non-existent correlation between an average sentiment measure derived by aggregating across all social media comments and the same offline tracking survey. We discuss the implications of our findings for practitioners considering social media as a potential research tool.

INTRODUCTION

With consumers increasingly engaged in online social media, companies have struggled with the question of how to integrate social media into their marketing strategy. Many have treated social media as just another channel with which to promote to and engage with customers. However, the effectiveness of such tactics is largely unknown, in part due to the lack of metrics with which to measure success. Other firms have used social media as a marketing research tool to gauge customer brand sentiment. This application of social media has given rise to a growing industry of “listening platforms” that monitor online conversations and measure the sentiment expressed through online social media (Hofer-Shall 2010).

This latter application of social media raises several questions for the marketing research community. In general, we currently have a limited understanding of the behavior related to social media. As a result, researchers have little guidance as to how they should interpret the volumes of comments posted online, leading firms to rely on simplified measures such as the total volume of posted comments or the average sentiment expressed across all posted comments. For example, researchers have monitored the number of tweets as a measure of engagement with products (Rui, Whinston and Winkler 2009) or events, such as the Super Bowl (NYTimes.com 2009) or the 2011 British Royal Wedding (LA Times 2011). Additionally, researchers have tracked aggregate measures of opinion expressed in the text of posted comments to assess viewer reactions to television shows (Kinon 2010) or to predict stock market performance (Bollen, Mao and Zeng 2010).

While these aggregate metrics may offer some guidance for marketers to assess customer engagement and overall sentiment toward the brand, there are several limitations that have prevented marketing researchers from integrating social media listening into their research

programs. First, because online environments allow for the posting of open ended and free form content, individuals have the flexibility to comment on anything they wish. As a result, individuals commenting on a given brand may focus their remarks on different attributes of the brand (e.g., customer service vs. reliability of the product itself) or different products in the brand's product portfolio. This stands in contrast to traditionally employed surveys in which researchers elicit responses pertaining to specific topics of interest. The implication for social media researchers is that simple metrics based on an aggregation across comments (e.g., average sentiment) can be problematic as they ignore established differences between product or attribute specific evaluations and general brand impressions (Dillon et al 2001).

Second, the venue (i.e., website domain) to which an individual posts may be related to the opinion posted. That is, some websites may systematically attract more positive (or more negative) individuals to post depending on a number of factors, such as the format of the venue, the nature of the audience, and various site specific dynamics. As a result, depending on the mix of websites represented in the sample, observed changes in aggregate metrics may simply reflect shifts in the composition of websites in the data sample rather than any underlying shifts in overall perceptions of the brand. Again, offline research methodologies typically control for such sampling biases while online sentiment measures routinely disregard them.

Finally, there is the issue of representativeness. In traditional offline research, much care is taken to identify individuals who would be representative of the population of interest. In contrast, individuals self-select when contributing online opinions (e.g., Ying, Feinberg and Wedel 2006). For example, if a brand were interested in measuring customer satisfaction, any offline methodology would start with drawing a systematic and unbiased sample from its customer base. In contrast, methodologies employing social media data tend to examine all

posts, regardless of whether they were contributed by customers or non-customers. As a consequence, posted opinions may not accurately reflect evaluations of the brand based on customers' experiences. Due to these limitations, brand sentiment metrics constructed from an aggregation of online comments are not necessarily comparable to the measures obtained from traditionally accepted offline methods.

However, the sentiment expressed in online social media can still inform marketing researchers. While there are several factors that influence online posted opinions, these comments are nonetheless affected by the contributor's overall sentiment toward the brand (Buschken, Otter and Allenby 2011). This underlying brand sentiment is distinct from attribute-specific evaluations (Dillon et al 2001), venue effects (Chen and Kirmani 2011), and other venue-specific dynamics (Moe and Schweidel 2011). Therefore, in an effort to derive a metric of underlying brand sentiment from social media conversations, we explicitly model and control for these latter factors and separate their effects from that of the underlying brand sentiment on posted opinions.

Specifically, we consider all posted comments pertaining to a target brand and code the sentiment expressed as negative, neutral or positive. We model posted sentiment as an ordered probit process and separate brand sentiment (a construct similar to the general brand impressions measure proposed by Dillon et al, 2001) from product- and attribute-specific evaluations. Additionally, we consider how the sentiment expressed in a comment varies depending on the website to which the comment is posted. We do so by allowing for both website-specific random effects and systematic effects across venue-types, which we define as the format of posting environment (e.g., blog, micro-blog, discussion forum, ratings and reviews, etc.). After controlling for these effects, we obtain a time-varying measure of brand sentiment that has been

adjusted to control for these factors. We further separate this measure into the sentiment of customers and that of non-customers.

To evaluate our approach, we compare the brand sentiment measure derived from our analysis of online comments to the results of a traditional offline brand tracking survey that was conducted during the same time period. While we find virtually no relationship between a simple average of the sentiment expressed in posted comments (i.e., ignoring the venue and content of posted comments) and the results of the offline survey (correlation = -0.002), the correlation between our adjusted measure of customers' brand sentiment and the offline survey is $.629$. These results demonstrate the potential for social media listening to be used as a marketing research tool if the various factors that influence posting behavior are carefully considered.

Our empirical results show that customers tend to be more critical than non-customers whose opinions are not necessarily related to any first-hand experience. When we consider differences across venue formats, we find that discussion forums which allow for more social interactions tend to attract more negative comments. In contrast, venues that facilitate the broadcasting of opinions (i.e., blogs and micro-blogs) tend to be more positive. These results have significant implications for social media researchers since these differences are typically ignored when constructing simple aggregate measures of online sentiment.

The remainder of the paper proceeds as follows. We first review research that conceptualizes the individual's posting decision and discuss how various factors influence posting decisions. Next, we describe the social media and survey data we employ in our analysis. We then detail our analysis of the social media data and our derivation of sentiment measures based on the data. We discuss the empirical findings and compare the estimated

measures based on the social media data to measures from the survey data. Finally, we conclude with a discussion of social media listening as a marketing research tool.

ONLINE SENTIMENT

While social media researchers have focused on the measurement of aggregate sentiment online, it is important to understand the individual behavior driving the decision to express an opinion. Therefore, in this section, we review some of the extant research that examines an individual's decision to post an online opinion and discuss factors that influence posted product opinions.

Moe and Schweidel (2011) propose that the posting decision consists of two component decisions, an incidence decision (*whether* to post) and an evaluation decision (*what* to post). Many researchers have focused exclusively on the incidence decision and examined the factors that influence the total volume of online word-of-mouth (Duan , Gu and Whinston 2008, Berger and Schwartz 2011). However, Moe and Schweidel (2011) propose that the incidence and evaluation decisions are inter-related and driven by (1) post-purchase product evaluations and (2) social dynamics in the posting environment. For example, they show that individuals with extremely negative or extremely positive product opinions are more likely to post an opinion online than individuals with moderate opinions, subject to the social dynamics present. Their results integrate the findings of offline research showing that individuals with extremely negative opinions are more likely to engage in word-of-mouth activity (Anderson 1998) with online studies showing a predominance of positive word-of-mouth (Chevalier and Mayzlin 2006).

However, the tendency toward expressing extreme opinions online does not necessarily prevent individuals holding moderate opinions toward the brand from entering the conversation.

As brands often represent a portfolio of products (Aaker and Keller 1990), individuals can express an extreme opinion toward a specific product in the brand's portfolio even if they hold moderate opinions toward the brand as a whole. Likewise, brands can be described by a multitude of attributes (Zeithaml 1988, Kirmani and Zeithaml 1993), and evaluations of specific attributes have been shown to be distinct from general brand impressions (Dillon et al 2001). As such, the opinions provided online may represent only the individual's evaluation on a particular product or attribute and not the underlying sentiment toward the brand.

Several studies have also shown that online opinions can be influenced by audience and venue effects. For example, Schlosser (2005) shows that posters moderate their online opinions in the face of a varied audience, a result consistent with offline studies of multiple audience effects (Fleming et al. 1990). Moe and Schweidel (2011) further demonstrate how, over time, such social dynamics can influence the evolution of opinion in an online environment. As such, since audiences and participants vary across websites, it is likely that the sentiment expressed will vary across social media sites and exhibit differing trends.

Furthermore, a few recent studies have shown that the consumer's choice of where to post is strategic and related to how they evaluate the product being discussed. For example, Chen and Kirmani (2011) show that when an individual's goal is to influence others, that individual will post negative messages in a homogenous venue and positive messages in a heterogeneous forum to more effectively persuade others. Muniz and O'Guinn (2001) show that individuals seek out forums most similar to them when their objective is to build or strengthen network ties. These studies demonstrate that a poster's evaluation decision is inter-related with their venue-choice decision.

However, despite the documented effects of venue choice on posted opinions, few researchers have controlled for the variation present across venues when constructing sentiment measures, revealing a potential limitation of the extant work on online opinions. Though some researchers have restricted their analysis to a single venue, such as individual newsgroups (Kozinets 2002), Twitter (Jansen et al. 2009), a retailer's product review environment (e.g., Chevalier and Mayzlin 2006, Moe and Trusov 2011) or a third-party review website (e.g., Duan, Gu and Whinston 2008), this approach does not account for the relationship that may exist between the chosen venue and posted evaluations. Should there exist systematic differences in brand sentiment between venues, any analysis of a single venue would confound venue-specific factors (including venue-specific dynamics) with derived measures of overall brand sentiment.

Finally, posted opinions can differ systematically across posters, specifically customers versus non-customers. Bird, Channon and Ehrenberg (1970) showed that brand perceptions can vary substantially across individuals with different usage frequency, where current customers hold different opinions from those of non-customers. In fact, both online and offline studies have shown that "experts" with greater knowledge and experience with a product are more critical and more likely to express a negative opinion when compared to non-experts (Amabile 1983, Schlosser 2005). If we assume that customers possess more expertise than non-customers because of their experience with the product (e.g., Anderson and Sullivan 1993, Kuksov and Xie 2010), we would expect brand sentiment expressed by customers to be more negative than those expressed by non-customers. While the direction of this difference is an empirical question, it does highlight the need to differentiate between customers and non-customers when deriving measures of brand sentiment from social media data

Not only do opinions held by customers and non-customers differ, but they also reference different sources of information. Zeithaml (1988) posited that individuals with direct experience with a brand are more likely to evaluate intrinsic product attributes while others are more likely to reference external cues such as advertising and word-of-mouth. As such, the textual content of comments may indicate whether the poster is likely a customer or a non-customer and therefore has implications for the sentiment expressed in online opinions.

The above discussion has outlined several sources of variation to consider when measuring brand sentiment via social media listening. Beyond the underlying sentiment toward the brand, extant research has suggested that an individual's posted opinion varies depending on (1) the focal product and/or attribute, (2) the venue to which the comment is contributed, and (3) and whether or not the poster is drawing on his own consumption experience. However, despite the systematic effects documented at the level of the individual poster, many popular social media metrics are based on simple aggregate metrics (e.g., total volume of comments, average sentiment, etc.) that mask covariate effects and yield a flawed brand sentiment measure. Therefore, in the next sections we discuss an approach to modeling social media data and provide an adjusted brand sentiment metric that tracks a carefully designed and implemented offline survey.

DATA

Our analysis involves two datasets pertaining to a single technology brand. The first dataset contains consumer comments posted online in a variety of venues. This data was provided by Converseon, a leading online social media listening platform that monitors and records online conversations. Converseon monitors a large sample of website domains and

identifies comments pertaining to a target brand. These comments are recorded and the textual content is coded for a random sample of comments. The resulting data set contains approximately 500 postings per month. The online data we use for this analysis spans a 15 month period from June 2009 to August 2010 and contain 7,565 posted comments. These comments were found across over 800 domains that support user generated content.

The textual content of the comments were individually coded by a team of analysts at Converseon. First, comments were coded for sentiment where each comment was identified as positive, negative or neutral. Second, comments were coded to distinguish between those that reference customers' direct experience with the brand and those that were based more on word-of-mouth or other sources of information. For example, comments that provided an anecdote based on personal experience with the brand was identified as a comment from a customer with direct experience. In contrast, comments that referred to a third party source of information (e.g., a link to a press release about the brand) were coded as general comments ("non-customer"). While this is not a perfect measure of the poster's experience with the brand, it does serve as a proxy to differentiate between customers and non-customers. Finally, the subject of each posted comment is identified. Since the brand represented in this dataset offers a large product portfolio, we identify the focal product of each comment. Additionally, the brand can be evaluated along a variety of brand attributes ranging from customer service and support to the technological reliability of its products. Thus, we also identify the focal attribute of each comment. In our data, we identify and distinguish between 140 unique products and 59 brand attributes.

From the domains present in our data sample, nine different venue types were identified. Table 1 describes each venue format, the number of posted comments they represent in our data, and the proportion of each venue’s postings that were from customers.¹

Venue Format	Illustrative Website	Frequency	Customer Postings
Discussion Forum	forums.adobe.com	2728	93%
Micro-blog	twitter.com	2333	37%
Blog	wordpress.com	2274	23%
Social Network	linkedin.com	155	40%
Mainstream News	cnbc.com	36	3%
Social News	digg.com	19	47%
Wiki	adobe.wikia.com	10	50%
Video	vimeo.com	6	0%
Review sites	epinions.com	4	25%

Table 1. Frequency of Venue Formats

We complement the online data set with a second dataset created from a traditional offline survey conducted by the brand. This survey was administered over the telephone to a sample of 1055 registered customers. The survey was conducted in 10 monthly waves from November 2009 through August 2010, which overlaps with the period during which our online data were collected. The online data, however, contains five additional months of data before the survey started. The survey measured customer the brand using seven separate questions (e.g., “What is your overall opinion about [brand]?” and “How likely would you be to recommend [brand] to a peer or colleague?”). A factor analysis conducted on the seven individual survey items revealed a single factor with an eigenvalue greater than 1 that explained 65% of the observed variance. Given the single factor on which the survey items load and the high pairwise correlations among the survey items (ranging from .44 to .83), we employ an average response across the seven items to represent our survey-based brand sentiment measure.

¹ Websites of discussion forums and wikis vary across companies and industries. To provide readers with an example of these domains while maintaining the confidentiality of our data provider, we list the forum and wiki pages for a different software company, Adobe Systems.

MODEL

Our modeling objective is to measure brand sentiment using the large volume of individual comments in our data sample. Based on the expressed opinions in these comments, we separately identify a latent brand sentiment measure for customer versus non-customer. Our approach provides an adjusted brand sentiment measure that controls the effects of the comment's focal product, focal attribute, and posting venue.

We model the opinion expressed in each comment using a hierarchical Bayesian ordered probit process. For comment i in our dataset, let Y_i denote the posted opinion such that $Y_i=1$ for negative posts, $Y_i=2$ for neutral posts and $Y_i=3$ for positive posts. To estimate the probability associated with the sentiment expressed, we specify $U_i = U_i^* + \varepsilon_i$, where U_i^* is determined by covariate and random effects, and ε_i is idiosyncratic error. We decompose U_i^* into a venue-specific brand sentiment construct, $VSent$, and comment-specific random effects that allow for variation among comments within a venue for a given month:

$$(1) \quad U_i^* = VSent_i + \pi_{p(i)} + \alpha_{a(i)}$$

where $p(i)$ denotes the focal product and $a(i)$ denotes the focal attribute of comment i . We account for heterogeneity across comments related to the focal product and attribute through π and α , with $\pi_p \sim N(0, \sigma_\pi^2)$ and $\alpha_a \sim N(0, \sigma_\alpha^2)$.

The term $VSent$ accounts for variation across the different domains (and consequently different venue formats) to which comments are contributed over time. We define $VSent$ to be a function of (1) the general brand impressions when the comment is posted, (2) time-invariant differences across venues, and (3) temporal variation that occurs within a particular venue format. First, general brand impression (GBI) is specified as a latent construct that varies over time but is invariant across venue formats. This construct provides the key metric with which to

compare the survey-based brand sentiment. Second, differences related to venues are decomposed into a random effect, ϕ , associated with the website domain, $d(i)$, and a fixed effect associated with the venue format, $v(i)$. This component of the model allows us to capture both systematic differences across venue formats and unobserved heterogeneity across the large sample of domains to which posters contribute comments. Finally, we allow for variation in the expressed sentiment over time to be specific to the venue format.² This allows for dynamics specific to some venue formats (e.g., social dynamics in ratings and review forums) to influence the venue-specific sentiment measure without necessarily affecting the general brand impression that is common across all venue formats.

We define $VSent$ separately for customers versus non-customers as follows:

$$(2) \quad VSent_i = \begin{cases} \beta_{v(i),1} + \delta_{d(i)} + \beta_{v(i),2} GBI_{t(i)} + \phi_{v(i),t(i)}, & Customer_i = 1 \\ \beta'_{v(i),1} + \delta_{d(i)} + \beta'_{v(i),2} GBI'_{t(i)} + \phi'_{v(i),t(i)}, & Customer_i = 0 \end{cases}$$

where $Customer_i$ indicates whether or not the i^{th} comment refers to a customer's experience with the brand's offerings. The coefficients $\beta_{\cdot 1}$ and $\beta'_{\cdot 1}$ serve as intercepts for customer and non-customer comments, respectively, that are specific to the venue format. This allows for expressed sentiment to systematically differ across venue formats. We account for heterogeneity across domains of the same venue format through the parameter δ , where $\delta_d \sim N(0, \sigma_\delta^2)$. To account for variation in the general brand impression (GBI) from month to month that is common to all venues, we assume that $GBI_t \sim N(0, \sigma_{G1}^2)$ and $GBI'_t \sim N(0, \sigma_{G2}^2)$. The coefficients $\beta_{\cdot 2}$ and $\beta'_{\cdot 2}$ allow for the magnitude with which GBI and GBI' affects $VSent$ to differ across venues. In other words, conversations in some venue formats may be more sensitive to changes in general brand impressions, while those in other venue formats may be more insulated.

² Due to the sparseness of data and the limited number of observations posted to many of the domains we observe, we assume that this variation is the same across domains of a given venue format.

For identification purposes, we assume that $\beta_{12}=1$ and $\beta_{12}'=1$. Thus, GBI and GBI' directly impact the sentiment expressed in venue $v=1$ (assumed to be blogs), and the coefficients β_{v2} and β_{v2}' reflect the extent to which the sentiment expressed in other venue formats compare. If one were to assume that $\beta_{v2}=1$ and $\beta_{v2}'=1$ for all v , this would suggest that all venues are equally responsive to changes in GBI and GBI' . Allowing the impact of GBI and GBI' to be specific to a venue format is consistent with Dillon et al. (2001), who allow general brand impressions to differentially affect brand ratings on different attributes.

As the coefficients $\beta_{.2}$ and $\beta_{.2}'$ are time-invariant, all temporal variation in this component of the model is explained by changes in GBI and GBI' . To allow for variation across venue formats in terms of how posted opinions evolve over time, we incorporate variables to capture venue-specific time trends (φ_{vt} and φ_{vt}') assume that $\varphi_{vt} \sim N(0, \sigma_{\varphi 1}^2)$ and $\varphi_{vt}' \sim N(0, \sigma_{\varphi 2}^2)$. This specification allows us to differentiate between general changes in underlying brand sentiment (captured by GBI and GBI') and more localized fluctuations in opinions that only manifest in certain venue formats. For example, a news event that fundamentally changes underlying brand sentiment will likely result in observable shifts in expressed sentiment across all venues and thus will be captured by GBI and GBI' . On the other hand, a localized event or dynamic specific to a particular venue format has limited implications for the general brand impression and thus will be captured solely through the venue-by-month interactions φ_{vt} and φ_{vt}' .

To complete our model specification, we assume that ε_i is drawn from a standard normal distribution. This corresponds to the following ordered probit probabilities:

$$(3) \quad \Pr(Y_i = r) = \begin{cases} \Phi(-U_i^*), & r = 1 \\ \Phi(\mu_{v(i)} - U_i^*) - \Phi(-U_i^*), & r = 2 \\ 1 - \Phi(\mu_{v(i)} - U_i^*), & r = 3 \end{cases}$$

where μ_v are the cutoff criteria (that $\mu_v > 0$) to which the latent sentiment, U^* , is compared. Note that the cutoffs are also specific to venue-type. While the venue-specific effects in equations (1) and (2) allow for positive or negative shifts in expressed sentiment across the different venue formats, the format-specific cutoffs allows for the mix of negative, neutral, and positive expressed sentiments to differ. For example, if a venue format is generally less negative, $VSent$ will be greater than other venues formats while a venue format that encourages more positive comments relative to neutral comments will be captured by differences in the format-specific cutoffs μ .

We estimate the model described in equations (1) – (3) using WinBUGS (<http://www.mrc-bsu.cam.ac.uk/bugs/>), which draws from the marginal posterior distributions of the parameters of interest using MCMC. Three independent chains were run for 10,000 iterations. We discarded the first 5,000 iterations of each chain as a burn-in. Convergence was assessed both visually and by Gelman and Rubin's F-test (1992).

EMPIRICAL RESULTS

Model Fit

We begin our discussion of results by first examining model fit and estimate a series of nested models to evaluate the value of each model component. We begin by estimating a model (Model 1) in which we only consider the month to month variation in the general brand impression, captured through GBI in equation (2), but ignore the remaining sources of variation. That is, in addition to ignoring factors related to the venue format (i.e., $\beta_{v1} = \beta_{v1}' = \beta$ and $\beta_{v2} = \beta_{v2}' = 1$ for all v), this baseline model assumes that $GBI = GBI'$. We then incorporate random effects associated with specific products, brand attributes and domains (π, α , and δ) in Model 2. Next, in

Model 3, we allow for differences in general brand impressions expressed by customers and non-customers, relaxing the assumption that $GBI = GBI'$. We further incorporate systematic differences across venue formats (β_v and β_v') in Model 4. Though the specification in Model 4 allows for differences across venue formats, it assumes that all temporal variation is explained by GBI and GBI' (i.e., $\varphi_{vt} = 0$ and $\varphi_{vt}' = 0$). We relax this restriction in Model 5 by allowing month to month shifts in sentiment to vary across venues formats.

We compare this set of models by examining DIC and aggregate hit rate in Table 2. We calculate the hit rate for each comment as the estimated probability of the observed sentiment, and the aggregate hit rate is the average hit rate across comments.

Model	Description	DIC	Hit rate
1	Monthly differences	14934	.404
2	+ random effects from product, brand attribute, and domain	13946	.455
3	+ customer vs. non-customer difference	13758	.460
4	+ venue main effect	13449	.473
5	+ venue x time interactions	13446	.476

Table 2. Model Performance

Comparing our baseline model specification (Model 1), which ignores all differences except for monthly variation, to the full model specification (Model 5), we see an 18% improvement in hit rate from .404 to .476. Coupled with the reduction in DIC, the improvement in hit rate provides support for the inclusion of these model components and highlights the variation in expressed sentiment due to factors not related to general brand impressions.

Based on the model comparison in Table 2, we focus the remainder of our discussion on the results of the fully specified model (Model 5). We start by examining the product, brand attribute and domain effects.

Variation in sentiment across focal products and attributes

An appealing characteristic of the model is its ability to quantify the differences in sentiment across the focal products and brand attributes. This allows the brand manager to isolate how each product in its brand portfolio and how each aspect of their product or service delivery contribute to posted online opinions. In this section, we demonstrate this functionality of the model and examine the posterior estimates for π and α for a selection of products and brand attributes.

For the 20 most popular products in the brand portfolio, Figure 1 provides the posterior estimates of π . These estimates, in effect, reflect how each product is evaluated relative to the overall brand. For this brand, 18 out of the top 20 products positively contribute to online sentiment. However, products O and T are viewed more negatively relative to the overall sentiment toward the brand. From a brand manager's perspective, these results serve as a red flag and may indicate that some intervention is necessary for these two products.

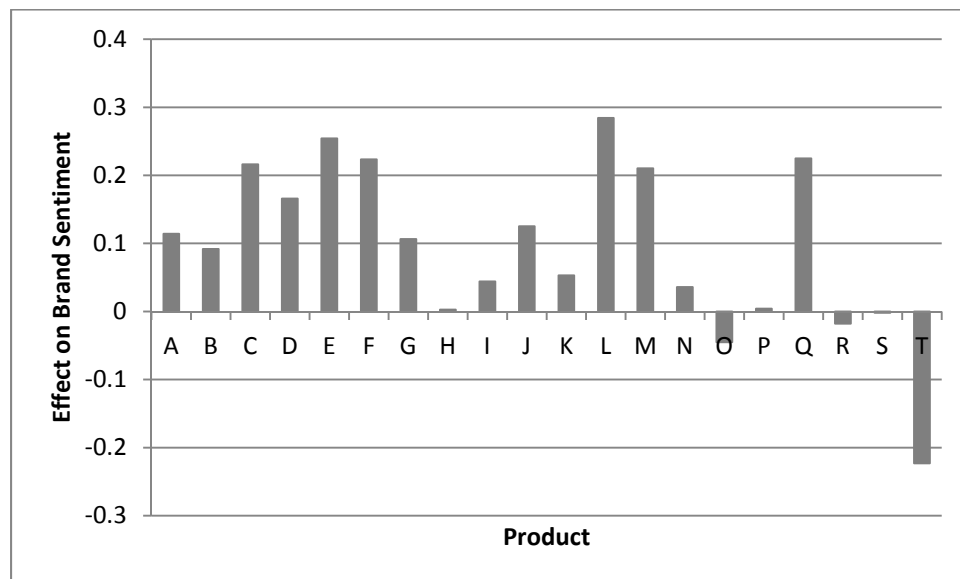


Figure 1. Effect of Focal Products on Brand Sentiment

Figure 2 illustrates the variation in sentiment across specific attributes relating to the brand. For illustration purposes, we provide the α estimates for 10 frequently mentioned attributes. While these findings are not generalizable to other brands, we provide these results to demonstrate the ability of our modeling approach to extract sentiment pertaining to specific aspects of the brand’s offerings. In this case, seven out of the ten frequently mentioned brand characteristics have a negative effect on expressed sentiment. The exceptions are brand reputation, quality of the product and size of the company. In other words, when posters focus on specific characteristics associated with product performance, their sentiment is more negative. In contrast, when reputation- and trust-related characteristics are evaluated, the sentiment expressed is more positive. For this brand, this indicates that while product functionality may receive critical comments online, the overall brand may be benefiting from a positive halo effect from past successes. Such a result may be cause for concern for the long-term future of the brand if the criticisms of product performance persist and continue to be discussed online.

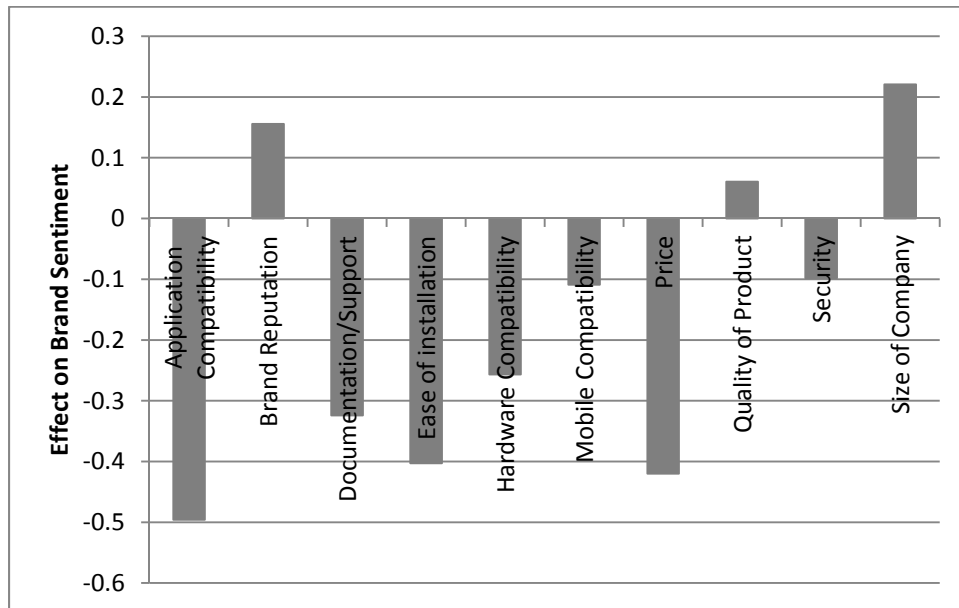


Figure 2. Effect of Focal Attribute on Brand Sentiment

Venue and customer differences

Venue effects on sentiment are captured by both domain specific random effects, δ , and systematic fixed effects associated with various venue formats, $\beta_{.1}$. In much the same way we illustrated product and attribute specific effects, we plot domain-specific effects for 10 frequently occurring domains in our dataset (Figure 3). The results indicate noticeable variation across this subset of domains. For example, the two social network domains represented in the figure have directionally opposite effects on how expressed sentiment deviates from the sentiment expressed in that venue type, where one website domain is attracting more positive opinions than the other. This result underscores the concerns associated with restricting an examination of online sentiment to a single domain.

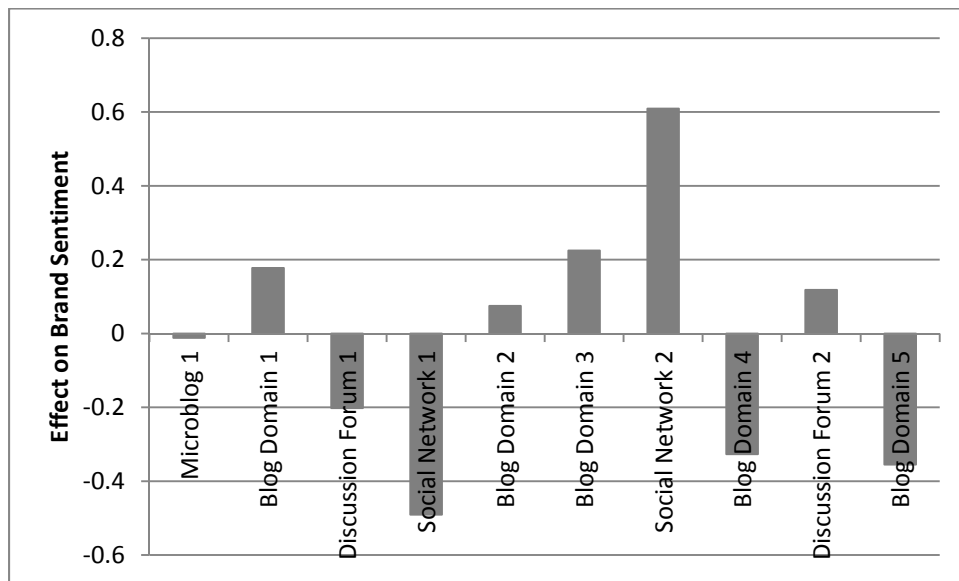


Figure 3. Differences in Brand Sentiment Across Domains

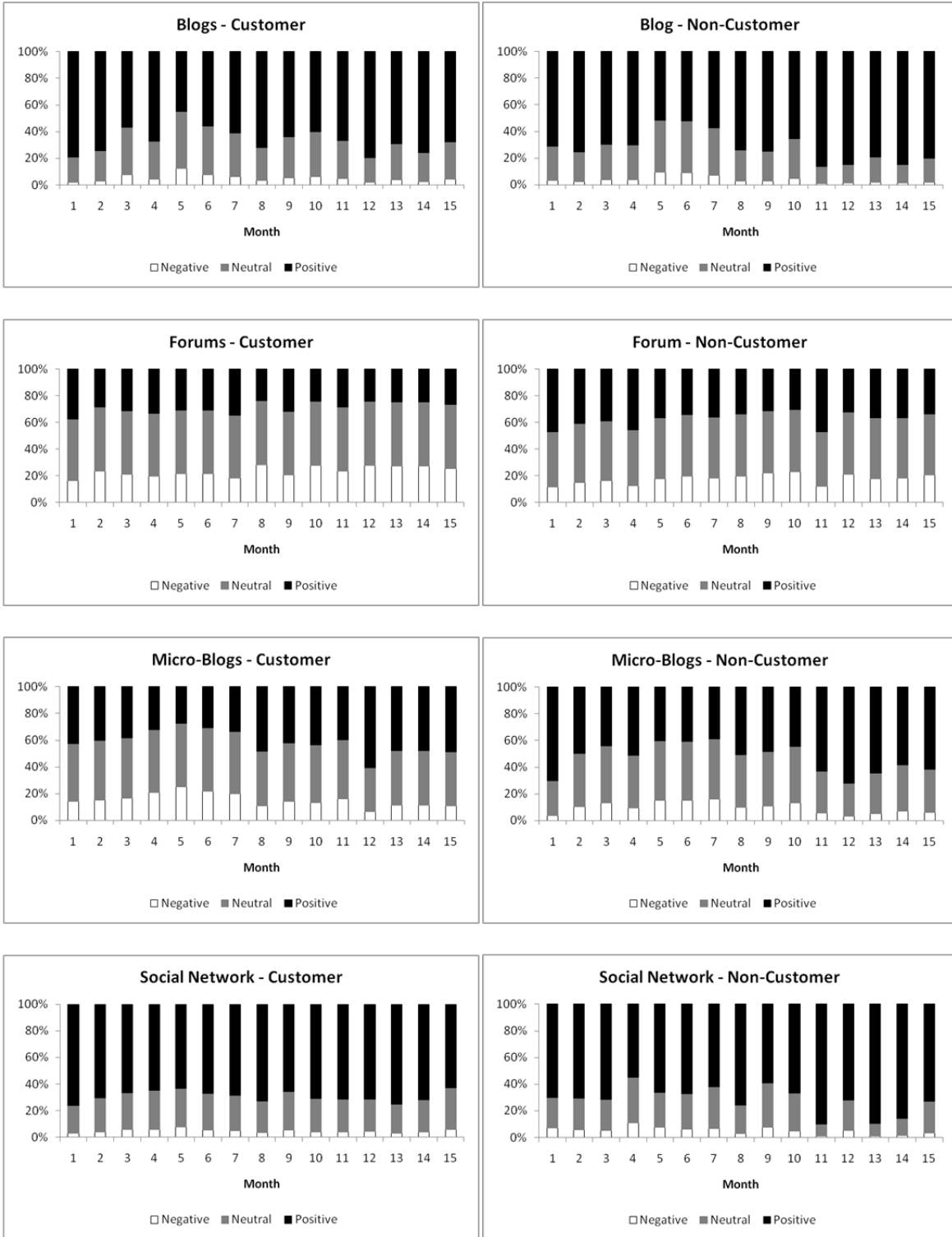


Figure 4: Expected Brand Sentiment Distribution across Venues

Differences across domains may result from a number of factors ranging from the design of the site, audience of readers, sponsor of the site, etc. Though these characteristics may be of interest to firms trying to expand their online presence to various social media sites, we treat these differences simply as random effects and focus instead on systematic differences that exist across venue formats. Figure 4 compares the expressed sentiment across different venue formats for customers and non-customers.

Across venue types, customer comments are more negative than non-customer comments. This result is consistent with previous research showing that “experts” tend to be more critical and provide more negative opinions (Amabile 1983, Schlosser 2005). Comparing across venue formats, we see that blogs and social networks are generally more positive than forums and micro-blogs. Due to differences in expressed sentiment across different venue formats, market researchers must be cognizant of the composition of their social media sample when constructing dashboard metrics. Neglecting differences across venue formats may result in the misleading inference that brand sentiment has shifted when the only change may be the proportion of different venue formats represented in the data.

To further illustrate the differences across venue formats, we focus the reader’s attention on only the sentiment associated with *customer* posts. Figure 5 plots the monthly means of the posterior distribution of $VSent$. We focus on the results from those venues that occur most frequently in our data and account for more than 99% of the observations in our data: blogs, forums, social networks, and micro-blogs. We also plot in Figure 5 the monthly mean of the posterior distribution for GBI_t , our adjusted measure of general brand impressions among customers. This measure captures the monthly variation common across all venue formats.

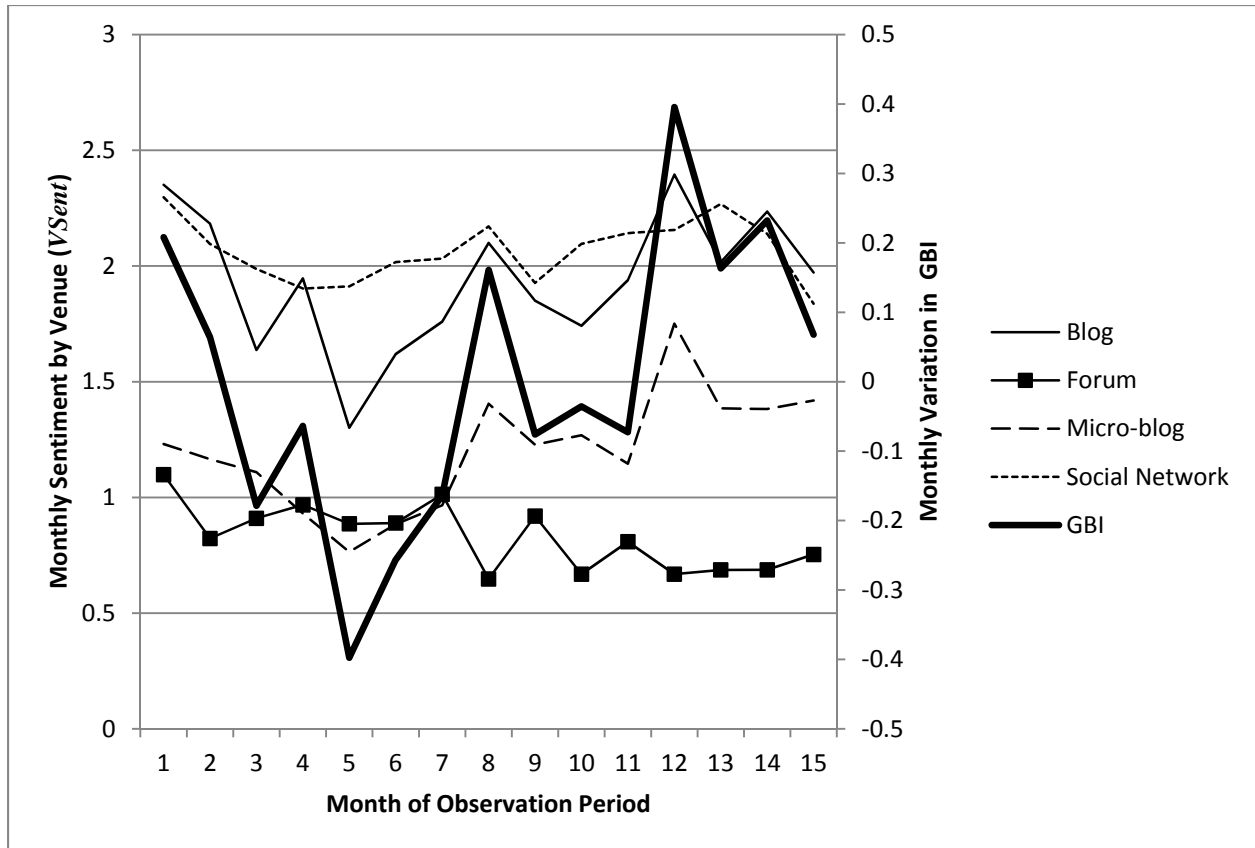


Figure 5. Sentiment Over Time

We find that the sentiment expressed in blogs and social networks track closely together through our observation period and is higher than the latent sentiment in micro-blogs and forums. Interestingly, while blogs, forums, and micro-blogs exhibit an upward trend after month 5 of the observation period, we find a slight downward trend in forums. This is consistent with prior research that show how social dynamics result in a negative trend in posted opinions (Moe and Schweidel 2011; Godes and Silva 2011).³

Overall, we see that our estimate of global brand impressions (*GBI*) generally tracks the venue-specific estimates of sentiment $V_{Sent_{vt}}$ for blogs, social networks and micro-blogs. For example, an industry-wide event in which the focal brand was unable to participate due to new guidelines implemented by the organizers (a competitor in the same industry) occurred in month

³ The public data collected from the social network sites more closely resemble micro-blogs among networked individuals. As a result, these comments do not feature the same degree of interactivity present in discussion forums.

4. The run-up to this event and the brand's inability to sponsor it may have contributed to the initial decline in sentiment we observe through month 5. After this decline, the adjusted measure of general brand impressions and the venue-specific sentiment measures for blogs, social networks and micro-blogs returned to previous levels and then stabilized. In contrast, the sentiment expressed in forums continues its slow, gradual decline. These differences across venue formats, as well as the departure of the venue-specific sentiment from the common monthly variation captured by *GBI*, highlight the need to account for venue-related differences in monitoring sentiment online. It remains a separate question, which we will take shortly, as to how closely *GBI* correlates with the variation in a traditional tracking survey.

Comparing the Adjusted Measure of Online Sentiment to an Offline Tracking Survey

The above results examine the various factors that may influence expressed sentiment. However, brand managers are less interested in the biases that exist online and more interested in the underlying brand sentiment that is common across venues, products, etc. Therefore, we compare next the brand sentiment measures derived from our model to the average sentiment obtained from a brand tracking survey.

Table 3 presents the correlations between reported customer satisfaction from the brand's offline tracking survey and various measures of online sentiment, all of which are derived from our modeling framework. When we compare the survey results to a simple average online sentiment, the correlation is virtually non-existent ($r=-.002$). This result should be very disconcerting for social media researchers that track online sentiment using aggregate summary statistics.

Measure	Correlation with Offline Survey
Observed online sentiment – overall average	-.002
Observed online sentiment – blogs	.197
Observed online sentiment – forums	.196
Observed online sentiment – micro-blogs	.394
Observed online sentiment – social networks	.394
<i>GBI</i>	.629

Table 3. Correlations Between Social Media Metrics and Offline Survey

In contrast, when we compare the offline survey results to our adjusted measure of brand sentiment (*GBI*), the correlation increases dramatically to .629. This comparison highlights the importance of controlling for venue effects, differences across poster experience and random effects for product and brand attributes when using social media as a listening tool.

We also examine the correlation between the tracking survey and observed sentiment expressed in the four venues that occur most frequently in our data (blogs, forums, micro-blogs and social networks). While these correlations are higher than the near zero correlation between the survey sentiment and average observed online sentiment, they are noticeably lower than the correlation between the survey and our adjusted brand sentiment measure. In other words, while aggregating observed sentiment across multiple venues provides a flawed measure of brand sentiment, measuring sentiment within a single venue does not entirely remedy the issue. Instead, leveraging the information across venues while controlling for factors that systematically influence expressed sentiment provides the best option for an online brand sentiment measure that tracks offline surveys.

DISCUSSION

In this research, we conduct a cross-venue analysis of sentiment as inferred from social media comments. In contrast to prior studies on social media that have focused on a single type of venue, our analysis reveals differences in the opinions expressed that exist across venues.

Moreover, these are not time-invariant systematic shifts. Rather, the sentiment expressed in different venues shift in distinct ways from month to month. We examine these venue effects for customers and non-customers separately.

After “backing out” deviations in sentiment that are specific to individual venues, we find that the monthly variation in brand sentiment that is common across venues closely relates to the offline tracking survey administered by the brand. This is in stark contrast to aggregate measures of observed opinions that are virtually uncorrelated with the survey results, demonstrating the value of the *GBI* measure in our modeling framework.

The proposed model also provides an approach for examining specific products in the brand portfolio or specific brand attributes, separately from the global brand. Consequently, social media listening may provide a powerful tool for brand managers interested in assessing individual elements of their brand.

Finally, the current research demonstrates the potential for social media to be incorporated into the brand’s marketing research activities. Listening to online social media conversations offers an economical and timely method from which brand sentiment can be inferred, suggesting that firms can supplement their current tracking studies with social media listening. These activities, however, must be undertaken with care. Monitoring a single type of venue would result in the inability to distinguish venue-specific factors from the general impressions of the brand. However, firms may be able to infer overall brand sentiment can be inferred from a broader sample of comments from multiple venues, provided that the variation across comments due to differences in the comments’ focal attributes and products, posting venue and customer experiences are carefully accommodated.

There are a number of directions that remain for future work. While we have accounted for differences in the venues to which social media comments are contributed, we have not investigated the specific characteristics of various social media venues that may influence expressed sentiment. Doing so may provide guidance to brands who are considering incorporating interactive components into their websites. Whereas our interest was in how closely brand sentiment inferred from social media mirrors tracking surveys, a popular marketing research technique, future research may also investigate the ability of changes in brand sentiment inferred from social media to predict shifts in key performance indicators such as market share or sales. While we demonstrate that the adjusted brand sentiment measure is more highly correlated with a tracking survey compared to the observed sentiment from any individual venue, specific venues may have superior predictive power. Finally, the current research demonstrates the potential for social media listening to supplement current research programs, but further investigation using data from a range of categories is essential before market researchers begin to rely exclusively on social media for customer insights.

REFERENCES

- Aaker, David A. and Kevin Lane Keller (1990), "Consumer Evaluations of Brand Extensions," *Journal of Marketing*, 54 (1), 27-41.
- Amabile, Teresa M. (1983), "Brilliant but Cruel: Perceptions of Negative Evaluators," *Journal of Experimental Social Psychology*, 19 (March), 146-56.
- Anderson, Eugene W. (1998), "Customer Satisfaction and Word of Mouth," *Journal of Service Research*, 1 (1), 5-17.
- Anderson, Eugene W. and Mary W. Sullivan (1993), "The Antecedents and Consequences of Customer Satisfaction for Firms," *Marketing Science*, 12 (2), 125-143.
- Berger, Jonah and Eric Schwartz (2011), "What Gets Talked About? Drivers of Immediate and Ongoing Word-of-Mouth," *Journal of Marketing Research*, forthcoming.
- Bird, M., C. Channon and A.S.C. Ehrenberg (1970), "Brand Image and Brand Usage," *Journal of Marketing Research*, 7 (3), 307-314.
- Bollen, Johan, Huina Mao and Xiao-Jun Zeng (2011), "Twitter Mood Predicts the Stock Market," *Journal of Computational Science*, 2 (1), 1-8.
- Buschken, Joachim, Thomas Otter and Greg Allenby (2011), "Do We Halo or Form? A Bayesian Mixture Model for Customer Satisfaction Data," *Working Paper*.
- Chen, Yu-Jen and Amna Kirmani (2011), "Persuading Others Online: The Consumer as Media Planner," working paper, University of Maryland.
- Chevalier, Judith and Dina Mayzlin (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 43 (3), 345-354.
- Dillon, William R., Thomas J. Madden, Amna Kirmani and Soument Mukherjee (2001), "Understanding What's in a Brand Rating: A Model for Assessing Brand and Attribute Effects," *Journal of Marketing Research*, 38 (4), 415-429.
- Duan, Wenjing, Bin Gu and Andrew B. Whinston (2008), "The Dynamics of Online Word-of-Mouth and Product Sales: An Empirical Investigation of the Movie Industry," *Journal of Retailing*, 84 (2), 233-242.
- Fleming, John H., John M. Darley, James L. Hilton, and Brian A. Kojetin (1990), "Multiple Audience Problem: A Strategic Communication Perspective on Social Perception," *Journal of Personality and Social Psychology*, 58 (April), 593-609.
- Gelman, Andrew and Donald B. Rubin (1992), "Inferences from Iterative Simulation Using Multiple Sequences," *Statistical Science*, 7 (4), 457-472.

Godes, David and Jose Silva (2011), "Sequential and Temporal Dynamics of Online Opinion," *Marketing Science*, forthcoming.

Hofer-Shall, Zach (2010), "The Forrester Wave: Listening Platforms, Q3 2010," Forrester Research.

Jansen, Bernard J., Mimi Zhang, Kate Sobel and Abdur Chowdury (2009), "Twitter Power: Tweets as Electronic Word of Mouth," *Journal of the American Society for Information Science and Technology*, 60 (11), 2169-2188.

Kinon, Cristina (2010), "Creators of 'Glee,' 'Modern Family,' track fan reaction via Twitter," NYDailynew.com, February 3, accessed at http://articles.nydailynews.com/2010-02-03/entertainment/27055284_1_creator-and-executive-producer-pregnancy-twitter.

Kirmani, Amna and Valarie Zeithaml (1993), "Advertising, Perceived Quality and Brand Image, in David A. Aaker and Alexander Biel, eds., *Brand Equity and Advertising*, Erlbaum: NJ, 143-162.

Kozinets, Robert V. (2002), "The Field Behind the Screen: Using Netnography for Marketing Research in Online Communities," *Journal of Marketing Research*, 39 (1), 61-72.

Kuksov, Dmitri and Ying Xie (2010), "Pricing, Frills, and Customer Ratings," *Marketing Science*, 29 (5), 925-943.

LA Times (2011), "Royal Wedding: Twitter, ABC News team up to track chatter online," April 28, accessed at [<http://latimesblogs.latimes.com/technology/2011/04/royal-wedding-abc-news-twitter-look-to-curate-chatter-on-prince-william-and-kate-middletons-big-day.html>].

Moe, Wendy W. and David A. Schweidel (2011), "Online Product Opinions: Incidence, Evaluation and Evolution," *Marketing Science*, forthcoming.

Moe, Wendy W. and Michael Trusov (2011), "The Value of Social Dynamics in Online Product Ratings Forums," *Journal of Marketing Research*, 48 (3), 444-456.

Muniz, Albert M., Jr. and Thomas C. O'Guinn (2001), "Brand Community," *Journal of Consumer Research*, 27 (4), 412-432.

NYTimes.com (2009), "Twitter Chatter During the Super Bowl," accessed at [http://www.nytimes.com/interactive/2009/02/02/sports/20090202_superbowl_twitter.html].

Rui, Huaxia, Andrew Whinston and Elizabeth Winkler (2009), "Follow the Tweets," *Sloan Management Review*, November 30, accessed at [<http://sloanreview.mit.edu/executive-adviser/2009-5/5152/follow-the-tweets/>].

Schlosser, Ann (2005), "Posting Versus Lurking: Communicating in a Multiple Audience

Context,” *Journal of Consumer Research*, Vol. 32, September, pp. 260-265.

Ying, Yuanping, Fred Feinberg and Michel Wedel (2006), “Leveraging Missing Ratings to Improve Online Recommendation Systems,” *Journal of Marketing Research*, 43 (3), 355-365.

Zeithaml, Valarie A. (1988), “Consumer Perceptions of Price, Quality, and Value: A Means-End Model and Synthesis of Evidence,” *Journal of Marketing*, 52 (3), 2-22.