Deciphering Word-of-Mouth in Social Media: Text-Based Metrics of Consumer Reviews

ZHU ZHANG, University of Arizona, Tucson
XIN LI, City University of Hong Kong
YUBO CHEN, University of Arizona, Tucson and Tsinghua University, Beijing, China

Enabled by Web 2.0 technologies, social media provide an unparalleled platform for consumers to share their product experiences and opinions through word-of-mouth (WOM) or consumer reviews. It has become increasingly important to understand how WOM content and metrics influence consumer purchases and product sales. By integrating marketing theories with text mining techniques, we propose a set of novel measures that focus on sentiment divergence in consumer product reviews. To test the validity of these metrics, we conduct an empirical study based on data from `Amazon.com` and `BN.com` (Barnes & Noble). The results demonstrate significant effects of our proposed measures on product sales. This effect is not fully captured by nontextual review measures such as numerical ratings. Furthermore, in capturing the sales effect of review content, our divergence metrics are shown to be superior to and more appropriate than some commonly used textual measures the literature. The findings provide important insights into the business impact of social media and user-generated content, an emerging problem in business intelligence research. From a managerial perspective, our results suggest that firms should pay special attention to textual content information when managing social media and, more importantly, focus on the right measures.

Categories and Subject Descriptors: I.2.7 [**Artificial Intelligence**]: Natural Language Processing—*Text analysis*; H.2.8 [**Database Management**]: Database Applications—*Data mining*

General Terms: Design, Performance

Additional Key Words and Phrases: Word-of-mouth, consumer reviews, text mining, sentiment analysis, social media

ACM Reference Format:
Zhang, Z., Li, X., and Chen, Y. 2012. Deciphering word-of-mouth in social media: Text-based metrics of consumer reviews. ACM Trans. Manag. Inform. Syst. 3, 1, Article 5 (April 2012), 23 pages.
DOI = 10.1145/2151163.2151168 http://doi.acm.org/10.1145/2151163.2151168

5

1. INTRODUCTION

With the advancement of information technology, social media have become an increasingly important communication channel for consumers and firms [Liu et al. 2010]. Social media are defined as a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0 and allow the creation and exchange of user-generated content (UGC) [Andreas and Michael 2010]. Social media can take many different forms, including Internet forums, message boards, product-review

This work was supported in part by CityU SRG-7002625 and CityU StUP 7200170.
Authors' addresses: Z. Zhang (corresponding author), Department of Management Information Systems, Eller College of Management, University of Arizona, Tucson, AZ; email: zhuzhang@eller.arizona.edu; X. Li, Department of Information Systems, City University of Hong Kong, Kowloon, Hong Kong, P.R.C.; email: Xin.Li@cityu.edu.hk; Y. Chen, Department of Marketing, Eller College of Management, University of Arizona, Tucson, AZ; email: yobochen@eller.arizona.edu.

DOI 10.1145/2151163.2151168 http://doi.acm.org/10.1145/2151163.2151168

websites, weblogs, wikis, and picture/video-sharing websites. Examples of social media applications include Wikipedia (reference), MySpace (social networking), Facebook (social networking), Last.fm (personal music), YouTube (social networking and video sharing), Second Life (virtual reality), Flickr (photo sharing), Twitter (social networking and microblogging), Epinions (review site), Digg (social news), and Upcoming.org (social event calendar).

Social media provide an unparalleled platform for consumers to share their product experiences and opinions, through word-of-mouth (WOM) or consumer reviews. With this platform, WOM is generated in unprecedented volume and at great speed, and it creates unprecedented impacts on firm strategies and consumer purchase behavior [Dellarocas 2003; Godes et al. 2005; Zhu and Zhang 2010]. Chen and Xie [2008] argue that one major function of consumer reviews is to work as sales assistants to provide matching information, to help customers find products matching their needs. This suggests that considerable value of consumer reviews, from a marketer's perspective, lies in the textual content.

However, when studying the sales impact of consumer reviews, most previous studies focus on the nontextual measures, such as numerical ratings (e.g., [Chevalier and Mayzlin 2006; Dellarocas et al. 2007; Duan et al. 2008; Zhu and Zhang 2010]. As shown in the literature [Herr et al. 1991; Mizerski 1982], some types of WOM information tend to be more diagnostic than others. Such nuance and diagnostic information is quite possibly "lost in translation" when the consumer is asked to provide only a numerical rating, which is an overly succinct summary opinion. An interesting analogy to this phenomenon is the academic refereeing process. While a numerical rating is provided as an overview, it is always preferable to have a detailed textual review to capture the subtleties of the evaluation. So far, how review content and metrics influence consumer purchases and product sales remain largely underexplored. This issue becomes even more important considering that most of WOM information in social media (e.g., message boards, blogs, chat rooms) does not come with numerical ratings [Godes and Mayzlin 2004; Liu 2006].

To address this important research issue, we propose novel content-based measures derived from statistical properties of textual product reviews. Based on the theoretical proposition that the role of consumer reviews is to provide additional information for customers to find matching products with their usage conditions [Chen and Xie 2008], we design the measures based on sentiment divergence contained in consumer reviews, which is independent from product attributes. To test the validity of such metrics from a marketer's perspective, we conduct an empirical study based on the book data from Amazon.com and BN.com (Barnes & Noble). Our results demonstrate a strong effect of our proposed measures on product sales, in addition to the effect of consumer rating metrics and other text metrics in the literature.

The rest of the article is organized as follows. Section 2 reviews previous literature and presents the theoretical background of this research. Section 3 develops our proposed sentiment divergence metrics for consumer reviews. Section 4 then presents the data, models, and findings of the empirical study, and Section 5 concludes with a summary and discussion of implications and future research opportunities.

2. LITERATURE REVIEW

In this section, we mainly review two bodies of related literature: (1) text sentiment mining, which is relevant to the computational basis of our metrics and (2) WOM as business intelligence for marketing and other management disciplines, which is related to the business dimension of this research.

2.1. Mining Sentiments in Text

Sentiment analysis has seen increasing attention from the computing community, mostly in the context of natural language processing [Pang and Lee 2008]. Studies in this area typically focus on automatically assessing opinions, evaluations, speculations, and emotions in free text.

Sentiment analysis tasks include determining the level of subjectivity and polarity in textual expressions. Subjectivity assessment is to distinguish subjective text units from objective ones [Wilson et al. 2004]. Given a subjective text unit, we can further address its polarity [Nasukawa and Yi 2003; Nigam and Hurst 2004; Pang et al. 2002; Turney 2001; Pang and Lee 2008] and intensity [Pang and Lee 2005; Thelwall et al. 2010]. Affect analysis, which is similar to polarity analysis, attempts to assess emotions such as happiness, sadness, anger, horror, etc. [Mishne 2005; Subasic and Huettner 2001]. In recent years, sentiment analysis has been combined with topic identification for more targeted and finer-grained assessment of multiple facets in evaluative texts [Chung 2009; Yang et al. 2010].

There are, in general, two approaches to text sentiment analysis. The first method is to first employ lexicons and predefined rules to tag sentiment levels of words/phrases [Mishne 2006] in text and then aggregate to larger textual units [Li and Wu 2010; Liu et al. 2005]. Linguists have compiled several lexical resources for sentiment analysis, such as SentiWordNet [Esuli and Sebastiani 2006]. Lexicons can be enriched by linguistic knowledge [Zhang et al. 2009; Subrahmanian and Reforgiato 2008] or derived statistically from text corpora [Yu and Hatzivassiloglou 2003; Turney 2001].

As compared with the lexicon-based approach, a learning-based approach aims at finding patterns from precoded text snippets using machine learning techniques [Dave et al. 2003], including probabilistic models [Hu and Li 2011; Liu et al. 2007], support vector machines [Pang et al. 2002; Airoldi et al. 2006], AdaBoost [Wilson et al. 2009], Markov Blankets [Airoldi et al. 2006], and so forth. To build effective models, various linguistic features (such as n-gram and POS tags) [Zhang et al. 2009] and feature selection techniques [Abbasi et al. 2008] have been used to capture subtle and indirect sentiment expressions in context and to align with application requirements [Wiebe et al. 2004]. There have been effective sentiment analysis tools, such as Opinion Finder [Wilson et al. 2005], developed based on previous research. In a learning-based approach, training classifiers generally requires manually coded data aligned with target applications at the word/phrase [Wilson et al. 2009], sentence [Boiy and Moens 2009], or snippet [Liu et al. 2007; Hu and Li 2011] level. Since manual coding of training data is typically labor-intensive, one can also first assess sentiments in smaller textual units with learning-based models and then aggregate to larger ones [Das and Chen 2007; Yu and Hatzivassiloglou 2003].

Table I summarizes major previous algorithmic efforts on sentiment analysis. We notice that a large number of studies took an aggregation approach based on either lexicons or statistical model outputs at a finer granularity. Importantly, most studies aim at assessing sentiment valence in text and characterizing the central tendency; there have been few that examine the distribution and divergence of sentiments in snippets or cross snippets. We are motivated to fill in this gap by proposing the sentiment divergence metrics.

Sentiment analysis has been applied to summarize people's opinions in news articles [Yi et al. 2003], political speeches [Thomas et al. 2006], and Web contents [Efron 2004]. Recent studies have extended sentiment and affect analysis to Web 2.0 contents, such as blogs and online forums [Liu et al. 2007; Li and Wu 2010]. In particular, due to their obvious opinionated nature, consumer reviews on products and services have received much attention from sentiment-mining researchers [Pang et al. 2002; Turney

Table I. A Brief Summary of Sentiment Analysis Methods

Study	Analysis Task	Sentiment Identification		Sentiment Aggregation		Nature of Measure
		Method	Level	Method	Level	
Hu and Li 2011	Polarity	ML (Probabilistic model)	Snippet			Valence
Li and Wu 2010	Polarity	Lexicon/Rule	Phrase	Sum	Snippet	Valence
Thelwall et al. 2010	Polarity	Lexicon/Rule	Sentence	Max & Min	Snippet	Range
Boiy and Moens 2009	Both	ML (Cascade ensemble)	Sentence			Valence
Chung 2009	Polarity	Lexicon	Phrase	Average	Sentence	Valence
Wilson et al. 2009	Both	ML (SVM, AdaBoost, Rule, etc.)	Phrase			Valence
Zhang et al. 2009	Polarity	Lexicon/Rule	Sentence	Weighted average	Snippet	Valence
Abbasi et al. 2008	Polarity	ML (GA + feature selection)	Snippet			Valence
Subrahmanian and Reforgiato 2008	Polarity	Lexicon/Rule	Phrase	Rule	Snippet	Valence
Tan and Zhang 2008	Polarity	ML (SVM, Winnow, NB, etc.)	Snippet			Valence
Airoldi et al. 2007	Polarity	ML (Markov Blanket)	Snippet			Valence
Das and Chen 2007	Polarity	ML (Bayesian, Discriminate, etc.)	Snippet	Average	Daily	Valence
Liu et al. 2007	Polarity	ML (PLSA)	Snippet			Valence
Kennedy and Inkpen 2006	Polarity	Lexicon/Rule, ML (SVM)	Phrase	Count	Snippet	Valence
Mishne 2006	Polarity	Lexicon	Phrase	Average	Snippet	Valence
Liu et al. 2005	Polarity	Lexicon/Rule	Phrase	Distribution	Object	Range
Mishne 2005	Polarity	ML (SVM)	Snippet			Valence
Popescu and Etzioni 2005	Polarity	Lexicon/Rule	Phrase			Valence
Efron 2004	Polarity	ML (SVN, NB)	Snippet			Valence
Wilson et al. 2004	Both	ML (SVM, AdaBoost, Rule, etc.)	Sentence			Valence
Nigam and Hurst 2004	Polarity	Lexicon/Rule	Chunk	Rule	Sentence	Valence
Dave et al. 2003	Polarity	ML (SVM, Rainbow, etc.)	Snippet			Valence
Nasukawa and Yi 2003	Polarity	Lexicon/Rule	Phrase	Rule	Sentence	Valence
Yi et al. 2003	Polarity	Lexicon/Rule	Phrase	Rule	Sentence	Valence
Yu and Hatzivassiloglou 2003	Both	ML (NB) + Lexicon/Rule	Phrase	Average	Sentence	Valence
Pang et al. 2002	Polarity	ML (SVM, MaxEnt, NB)	Snippet			Valence
Subasic and Huettner 2001	Polarity	Lexicon/Fuzzy logic	Phrase	Average	Snippet	Valence
Turney 2001	Polarity	Lexicon/Rule	Phrase	Average	Snippet	Valence

(Both = Subjectivity and Polarity; ML = Machine Learning; Lexicon/Rule = Lexicon enhanced by linguistic rules).

2001]. Several studies have evaluated sentiment analysis methods on movie or product review corpora [Hu and Li 2011; Kennedy and Inkpen 2006]. There also have been system-driven efforts to include sentiment analysis into purchase decision marking. For instance, Popescu and Etzioni [2005] introduce an unsupervised system to extract important product features for potential buyers. Liu et al. [2005] propose a framework that compares consumer opinions on competing products. These efforts can serve as preliminary computational infrastructure for social media-driven business intelligence, which will be reviewed in the next section.

2.2. WOM as Business Intelligence: Textual Metrics

The emergence of social media promotes collective intelligence among online users [Surowiecki 2005], which affects individuals' decisions and influences organizations' operations. WOM in social media has started to gain much attention from business managers as an emerging type of business intelligence [Chen 2010].

In the finance literature, for example, social media have been used as indicators of public/investor perception of the stock market. Through manually extracting the "whisper forecasts" in forum discussions, Bagnoli et al. [1999] find that unofficial forecasts are actually more accurate than professional analysts' forecasts in predicting stock trends. Tumarkin and Whitelaw [2001] find that days of abnormally high message activity in stock forums coincide with that of abnormally high trading volumes, when online opinion is also correlated with abnormal returns. After applying the Naive Bayes algorithm to classify the sentiments of stock-related forum messages into three rating categories (bullish, bearish, or neither), Antweiler and Frank [2004] find that the overall sentiments help predict market volatility and that the variance of the rating categories is associated with increased trading volume. Das et al. [2005] and Das and Chen [2007] introduce more statistical and heuristic methods to assess forum message (and news) sentiments. They show that both message volume and message sentiment are significantly correlated with stock price, but a coarse measure of disagreement is not significantly correlated with either market volatility or stock price. Gao et al. [2006] study the effect of a numerical divergence measure (market return volatility) on IPO performance and find a negative correlation between the two.

Recently, several published marketing studies have examined how WOM in social media influences product sales (e.g., [Chen et al. 2011; Chevalier and Mayzlin 2006; Godes and Mayzlin 2004; Liu 2006; Zhang 2008; Zhu and Zhang 2010]), which is very important for marketing managers. Some of the most commonly studied WOM measures in this body of literature include volume (i.e., the amount of WOM information) and valence (i.e., overall sentiments of WOM, positive or negative). For instance, Liu [2006] shows the volume of WOM has an explanatory power for movie box office revenue. Differently, Godes and Mayzlin [2004] find the dispersion of conversation volumes across communities (instead of the volume itself) has an explanatory power in TV viewership. Chevalier and Mayzlin [2006] demonstrate that the valence of consumer ratings has a significant impact on book sales at `Amazon.com`. Zhang [2008] demonstrates a positive correlation of usefulness-weighted review valence with product sales across four product categories. In addition, Forman et al. [2008] find that the disclosure of reviewer identity information and a shared geographical location between the reviewer and customer increases online product sales, therefore suggesting the potential impact of some interesting factors beyond the reviews themselves.

However, almost none of these measures are constructed directly from the textual content of consumer reviews. Some recent studies made attempts close to the neighborhood, but all have pitfalls. Liu [2006] studies the sales effect of review sentiment valence based on human coding of the review text. Liu et al. [2010] examine review sentiment valence and subjectivity with text-mining techniques. In both studies, none of these content-based review measures is found to have an impact on or a correlation with product sales. Ghose and Ipeirotis (forthcoming) find significant effect of review subjectivity on the sales of audio-video players but not on digital cameras or DVDs.

3. SENTIMENT DIVERGENCE METRICS OF CONSUMER REVIEWS

The preceding literature review has revealed that the community is increasingly paying attention to quantitative measures of consumer-generated product reviews and their influence on product sales. On the other hand, very limited research examines content-based textual metrics of consumer reviews. The extant early attempts in this area mainly focused on the sentiment valence but failed to find a significant sales effect of such textual metrics (e.g., Liu et al. [2010]).

In light of the limitations of previous research, in this article, we propose a set of sentiment divergence measures of consumer reviews and investigate how they affect product sales, in addition to all existing nontextual and textual valence review metrics. As compared with previous measures, (1) our metrics are designed to capture text-based opinion divergence (instead of valence) in product reviews, which adventure into an unexplored frontier in sentiment analysis, and (2) we focus on their impact on consumer purchase behavior and product sales, which has significant implications for WOM-driven marketing research.

In this section, we first propose our sentiment divergence metrics and then develop hypotheses on how these metrics affect consumer purchases and product sales.

3.1. Sentiment Divergence Metrics

To facilitate the discussion, we assume that we have a collection of products P = $\{p_1, \ldots, p_i, \ldots, p_n\}$, and each product pi is rated and reviewed by Ci customers. Therefore, each pi is associated with two sets of information.

—A rating set $R_i = \{r_i{}^1, \ldots, r_i{}^j \ldots, r_i{}^{Ci}\}$, where each $r_i{}^j$ is a positive real number, typically between 1 and 5, as in many e-commerce websites.
—A review set $V_i = \{v_i{}^1, \ldots, v_i{}^j \ldots, v_i{}^{Ci}\}$, where each $v_i{}^j$ is a piece of evaluative text reflecting a consumer's opinions on the given product, which can be null if the consumer chooses not to provide it. Such product reviews are, again, widely available on many e-commerce websites.

The items in the two sets correspond to each other pairwise, that is, each rating $r_i{}^j$ corresponds to a review $v_i{}^j$.

In order to define the sentiment divergence metrics, we first quantify the sentiments associated with individual words in the product reviews. Generally, a word can be either objective or subjective. A subjective word can carry either a positive or a negative sentiment with certain intensity. For example, "fantastic" is a strongly positive word; "questionable" is a moderately negative word; and "white" is a neutral word. We take a lexicon-based approach and use SentiWordNet [Esuli and Sebastiani 2006] to code word sentiments. In this lexicon, an English word w is associated with two scores, a positivity score *pscore* and a negativity score *nscore*, where $0 \leq pscore, nscore \leq 1$, and $0 \leq pscore + nscore \leq 1$. A word, in most circumstances, carries a positive sentiment if its *psocre* is significantly larger than its *nscore*, and vice versa specifically, we combine the *pscore* and *nscore* to a microstate score *ms* for each word to indicate positive, negative, and neutral words,[1]

—$ms = 0$, if the word is neutral ($|pscore\ nscore| < 0.1$).
—$ms = 1$, if the word is positive ($pscore\ nscore \geq 0.1$).
—$ms = -1$, if the word is negative ($nscore\ pscore \geq 0.1$).

Figure 1 illustrates the process of converting a sentence first to a word vector, then to a polarity score vector (PSV), and last to a microstate sequence.

Based on this representation of text, we can approximately depict the underlying sentiment distribution of review content. Specifically, we design two measures to capture the diversity of the opinions delivered in all reviews for each product, which may possibly influence customers' purchase decisions.

The mathematical basis of our measures is the Kullback-Leibler divergence [Kullback and Leibler 1951] developed in information theory, which we use to measure the difference among multiple microstate sequences (i.e., reviews) associated with a given product. The Kullback-Leibler divergence is an asymmetric measure of the difference

[1]SentiWordNet provides sentiment scores for about 115,000 word entries. The difference between *pscore* and *nscore* approximately characterizes the polarity (neutral, positive, or negative) of the corresponding word entry. Neutral words typically have 0 values on both scores with a few exceptions that have nonzero but similar values (e.g., the word "heavy" has a pscore of 0.125 and nscore of 0.125). 0.1 is an appropriate threshold that partitions the entire word space into three regions with the neutral words (about 88,500) in the middle region and the positive/negative words in the two tails.

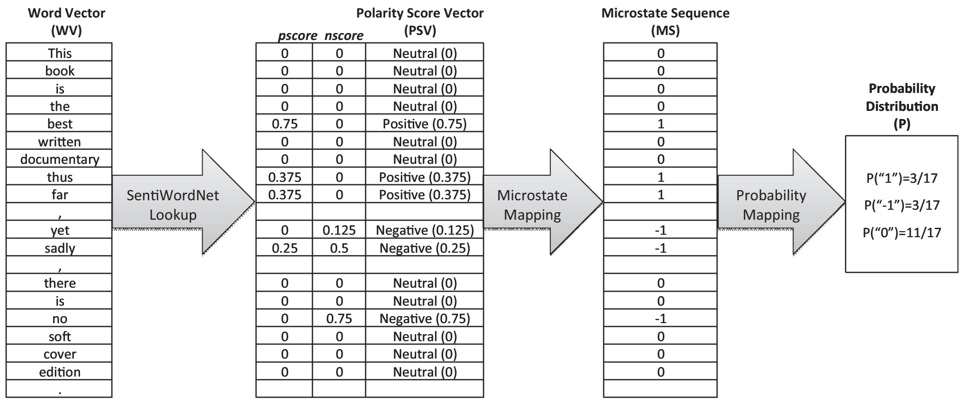

Fig. 1. Conversion of text representation.

between two probability distributions. Given two reviews v_1 and v_2 which are mapped into two microstate sequences MS_1 and MS_2, we can calculate their probability distributions P_1 and P_2 according to the *ms* scores. The Kullback-Leibler divergence is defined as

$$KL(P_1 \parallel P_2) = -\sum_x p_1(x) \log p_2(x) + \sum_x p_1(x) \log p_1(x)$$

Since in this research we only care about the difference between comments and ignore their ordering, we use a symmetrized Kullback-Leibler divergence measure, which is defined based on the Kullback-Leibler divergence as

$$SymKL(P_1, P_2) = KL(P_1 \parallel P_2) + KL(P_2 \parallel P_1).$$

Based on these notations, we propose the following two *sentiment divergence* measures.

Sentiment KL Divergence (SentiDvg_KL). The first measure is defined as the average pairwise symmetrized Kullback-Leibler divergence between all reviews v_i^j in V_i such that

$$SentiDvg\_KL_i = \frac{\sum_{j=1}^{C_i} \sum_{k=1,k\neq j}^{C_i} SymKL\left(P_i^j, P_i^k\right)}{C_i(C_i - 1)}.$$

Sentiment JS Divergence (SentiDvg_JS). The second divergence measure is based on the Generalized Jensen-Shannon Divergence [Lin 1991]. Instead of conducting pairwise comparison between reviews (as in *SentiDvg_KL*), this measure compares each review with an artificial "average" review. Formally, based on the microstate sequences corresponding to a product, we first calculate the average probability distribution of all P_i^j's as *AvgP*. The *SentiDvg_JS* is then defined as

$$SentiDvg\_JS_i = \frac{\sum_{j=1}^{C_i} SymKL\left(P_i^j, AvgP\right)}{C_i}.$$

The two measures are designed with the same motivation and are not meant to be qualitatively different. Intuitively, given a *population* of reviews of a product, SentiDvg_KL measures the total pairwise dispersion between each other, and SentiDvg_JS quantifies the total dispersion between every individual and a hypothetical *center*. They may behave somewhat differently depending on empirical data.

3.2. Sentiment Divergence and Consumer Purchase Behavior: Theories and Hypotheses

Our proposed sentiment divergence metrics capture the distribution of sentiments in product reviews. In this research, we investigate how they are different from existing measures and how they affect product sales.

3.2.1. Sales Effects of Review Sentiment Divergence. There are at least two streams of literature that shed light on the potential sales effects of review sentiment divergence.

The Function of Word of Mouth and Consumer Reviews. Chen and Xie [2008] argue that one major function of consumer reviews is to work as a new element of the marketing communication mix to provide product-matching information (mainly available in review text) for customers to find products matching their usage conditions. In product markets, very few products match the preferences of all consumers. For any product, there always exist some matched and some unmatched customers [Hotelling 1929]. Meanwhile, consumer reviews largely reflect subjective evaluations based on preference matches with the product [Li and Hitt 2008]. In fact, Liu [2006] suggests that one main reason he failed to find the significant sales impact of sentiment valence of consumer reviews lies in such preference heterogeneity. Negative reviews due to product mismatch might be quite useful and reduce the uncertainty for potential matching. As a result, the informativeness of consumer reviews may be largely reflected by the degree of opinion divergence in the review content. An overly convergent set of opinions might provide a smaller amount of product-matching information, which leads to a smaller group of converted consumers.

The Quality of Product Information. The extant information marketing literature argues that the quality of the information in a market depends on both the reliability of an information source and the correlation among information sources (e.g., Sarvey and Parker [1997]). When the information is less reliable, combining multiple information sources less positively correlated will provide higher information accuracy. Similarly, recent economics studies on news markets argue that a reader who aggregates news from different sources can get a more accurate picture of reality, particularly when the news sources are highly diversified [Mullainathan and Shleifer 2005; Gentzkow and Shapiro 2006]. For a given product, each consumer review (as an individual information source) is less reliable, given that it is from an anonymous consumer instead of experts with trustworthy identities. Therefore, when evaluating a product via reading different consumer reviews, more divergent (thus less correlated) opinions will lead to more a informed purchase decision. Given the overwhelming number of positive reviews [Chevalier and Mayzlin 2006], such increased persuasion power will lead to increased sales.

In light of these theories, product reviews with high sentiment divergence may provide more product-matching information with higher reliability. When consumer sentiments are more divergent, the posted consumer reviews tend to reflect preferences of more heterogeneous consumers and thus provide more matching information between product attributes and usage situations for a larger population of potential buyers. Furthermore, consumers reading divergent product reviews would be able to combine them for a more comprehensive assessment of products. Thus, we expect that the sentiment divergence measures have a positive correlation with product sales, hence the following hypothesis regarding the sales impacts of sentiment divergence of consumer reviews.

H1. Consumer reviews with higher sentiment divergence have larger impacts on sales than those with lower sentiment divergence.

3.2.2. Sentiment Divergence vs. Rating Divergence. Notice that along with consumer reviews, most e-commerce websites also provide numerical ratings associated with review texts, $R_i = \{r_i{}^1, \ldots, r_i{}^j \ldots, r_i{}^{Ci}\}$. It is important to point out that the information captured by our sentiment divergence is at the textual content level and may not be captured by ratings. Furthermore, the rating itself does not carry product-matching information. Therefore, textual review sentiment divergence provides additional information and generates sales impacts in addition to the effects of consumer ratings.

Similar to sentiment divergence, consumer ratings also carry variance or disagreement information [Sun 2009]. However, this information influences consumer purchase decisions differently. Research in consumer behavior and decision making has shown that the variance or disagreement among critic ratings can create uncertainty for consumers [Kahn and Meyer 1991; West and Broniarczk 1998]. Based on the prospect theory [Kahneman and Tversky 1979], West and Broniarczk [1998] show that consumers' responses to uncertainty in the form of critic disagreement depends on their expectations. In situations in which consumers have high expectations on product ratings, they will frame the product evaluation task as a loss or as utility-preserving. The critic disagreement increases the chance of meeting or exceeding their goals, and consumers will have higher evaluations when there is critic disagreement than when there is agreement. In contrast, when expectations are low, consumers will frame the product evaluation task as a gain or as utility-enhancing. The critic disagreement increases the possibility of falling short of their expectations, and therefore consumers will have higher evaluations when there is critic consensus rather than disagreement. Therefore, the variance of consumer ratings will have a positive effect on consumer purchases when consumers have high expectations on ratings, but it will have a negative sales effect when they have low expectations on ratings.

Researchers have found that, due to self-selection bias, a majority of consumer reviews are positive in the earlier time period but decrease in positivity over time [Dellarocas et al. 2007; Li and Hitt 2008; Zhu and Zhang 2010]. This suggests that consumers might have high expectations on consumer ratings in the earlier time period and low expectations in the later time period. Therefore, we would expect that the variance of consumer ratings will have different impacts on consumer purchases over time. We might observe a positive sales effect of consumer rating variance earlier but a negative effect over time. On the other hand, the main function of consumer textual review content is to provide product-matching information [Chen and Xie 2008]. Sentiment divergence based on review texts mainly plays an informative role, which remains positive and largely unchanged over time.

We have the following hypotheses regarding the unique impact of sentiment divergence of consumer reviews from the variance of consumer ratings over time.

H2a. The impact of the variance of consumer rating on sales will be likely to shift from positive to negative over time.

H2b. The impact of review sentiment divergence on sales will be less likely to shift and will remain positive over time.

4. EMPIRICAL STUDY

In this section, we collect a longitudinal dataset from `Amazon.com` and `BN.com`, and empirically validate our proposed textual metrics of consumer reviews.

In order for our proposed textual metrics of consumer reviews to be valid, they must be able to demonstrate a sales impact, after controlling for exiting review measures in the literature. Deploying a Difference-in-Difference (DID) approach, the seminal study by Chevalier and Mayzlin [2006] demonstrates a causal relationship between consumer ratings and product sales and constitutes an important benchmark for our study. We conduct our empirical study in the same setting, and we also focus on the book category and two major vendors in the book retail industry, `Amazon.com` and `BN.com`.

4.1. Data

Our data are collected from `Amazon.com` and `BN.com` during three time periods: a two-day period in the middle of November 2009, a two-day period in the middle of January 2010, and a two-day period in the middle of July 2010. Both websites have a large consumer base and a great number of consumer reviews. We collect data on all the books in the "Multicultural" Subcategory of Amazon's "Romance" category through Amazon Web Services (`http://aws.amazon.com`). We then conduct an ISBN search on `BN.com` to collect the matching titles if they are available.

For each book in our datasets, we gather the available product information, including sales rank, lowest price charged for the book (by either the website or third party sellers), ratings (on a scale of one to five stars, with five stars being the best), and textual content of all product reviews.[2] On both sites, the products are ranked according to the number of sales made in a time period. The top-selling product has a sales rank of one, and the less popular products are assigned higher sequential ranks.

We examine differences in sales over the November 2009–January 2010 horizon and over the November 2009–July 2010 horizon. The two different time spans are used to test the different predictions on consumer rating variance and sentiment divergence in Hypothesis 2 (H2a and b) and to show the robustness of our results on sentiment divergence in Hypothesis 1 (H1). As in previous studies (e.g., [Chen et al. 2011; Chevalier and Goolsbee 2003; Chevalier and Mayzlin 2006]), sales rank is used as a proxy for sales, given the linear relationship between ln(sales) and ln(sales rank).[3] We included in our final data only books that are available and have sales ranks at both `Amazon.com` and `BN.com` and that have more than one review at each website for all three periods. Our final sample includes 345 books, and Table II presents the summary statistics of the data. As in Chevalier and Mayzlin [2006], for each book there are many more consumer reviews on `Amazon.com` than on `BN.com`. Consistent with the literature (e.g., [Chevalier and Mayzlin 2006; Li and Hitt 2008; Zhu and Zhang 2010]), the average rating decreases over time, but the standard deviation (variance) of ratings increases. The sales ranks, in general, increase (i.e., sales decline) over time on each site.[4]

4.2. Models and Results

4.2.1. Model Specifications: A Difference-in-Difference (DID) Approach. The biggest challenge to demonstrating the sales effects of consumer reviews is the endogeneity issue [Chen et al. 2011; Chevalier and Mayzlin 2006]. Specifically, it is necessary to show that the identified correlations between the consumer review metrics and product sales are not spurious, and do not result from some unobserved factors. For instance, some

[2]Data on shipping policy for each book do not vary much during the two-month period of our data collection, and are thus not included here.

[3]Notice that sales rank of Amazon data may reflect sales up to a month [Chevalier and Mayzlin 2006]. To eliminate the possible interference between cooccurring product reviews and product sales, following Chevalier and Mayzlin [2006], there is a one-month lag between the review data and sales data in our model.

[4]The sales rank in January 2010 on Amazon is significantly higher even compared with that in July 2010. One explanation might be the holiday shopping at `Amazon.com`. Since that sales rank at Amazon may reflect sales up to a month [Chevalier and Mayzlin 2006], the data we collected covers the sales during Christmas. Consumers at `Amazon.com` tend to do more holiday shopping since they buy many holiday products (e.g., electronics) together with other popular books for Christmas gifts (e.g., children books). The sales rank on books in the Multicultural category at `Amazon.com` may be deflated by those popular books for Christmas gifts.

Table II. Summary Statistics

	Nov 2009		Jan 2010		Jul 2010	
	Amazon	BN	Amazon	BN	Amazon	BN
Sales Rank	544,350.6	263,774.4	604,288.9	305,579.8	588,216.7	329,305.0
	(426,261.1)	(204,411.0)	(466,434.2)	(218,433.0)	(478,119.9)	(230,583.3)
Price	4.21	5.24	4.43	5.49	4.40	5.52
	(2.74)	(2.83)	(2.88)	(2.90)	(2.79)	(3.32)
# Reviews per book	31.77	9.59	32.26	9.67	33.32	10.05
	(45.13)	(16.85)	(45.17)	(16.82)	(45.37)	(17.12)
Average Rating	4.28	4.47	4.27	4.47	4.26	4.45
	(.46)	(.51)	(.46)	(.51)	(.45)	(.52)
Std. Dev. of Rating	.857	.547	.870	.553	.885	.578
	(.362)	(.438)	(.355)	(.434)	(.340)	(.441)
SentiValence	.0871	.1046	.0870	.1047	.0877	.1049
	(.0241)	(.0588)	(.0239)	(.0589)	(.0231)	(.0579)
SentiDvg_KL	.1043	.2277	.1064	.2278	.1070	.2251
	(.0720)	(.3513)	(.0808)	(.3495)	(.0752)	(.3470)
SentiDvg_JS	.0263	.0554	.0273	.0557	.0276	.0552
	(.0215)	(.0839)	(.0248)	(.0837)	(.0231)	(.0825)
# Observation	345	345	345	345	345	345

Note: the number on the upper part of each cell is the mean of the variable, and the number in parenthesis is the corresponding standard deviation.

unobserved product characteristics, such as book contents, might lead to high review sentiment divergence and also high sales, which should not be mistaken as sentiment divergence's effect on product sales. One effective econometric method for addressing this endogeneity issue is the DID approach [Wooldridge 2002]. Chevalier and Mayzlin [2006] have used this approach to demonstrate the sales effects of consumer reviews on relative sales of books at `Amazon.com` and `BN.com`. In our article, we use their model as one of the benchmark models to demonstrate the initial value of our sentiment divergence metrics. A practical advantage of our method is incorporating the competitor information (`BN.com` relative to `Amazon.com`) to control for the confounding from product site effects (e.g., consumers at `Amazon.com` may prefer certain books more than users at `BN.com`, instead of the WOM effect on those books). In the following, we first briefly derive benchmark DID models and then introduce other related models with sentiment divergence metrics.

When making purchase decisions, consumers might read reviews and compare prices from both websites (`Amazon.com` or `BN.com`). Consider such potential competing effects: a book's sales rank on a website is dependent on its prices and consumer reviews in both sites. It also depends on unobserved book fixed effects ψ_i, such as book quality and promotion, and unobserved book-site fixed effects μ_i^A or μ_i^B for site A or B, respectively (e.g., consumers at `Amazon.com` may prefer certain books more than users at `BN.com`). Therefore, the sales rank of a book i at a site A (`Amazon.com`) or B (`BN.com`) at time t is

$$\ln(\mathrm{Rank}_{\mathrm{it}}^{\mathrm{A}}) = \omega_{0\mathrm{t}}^{\mathrm{A}} + \omega_{1\mathrm{t}}^{\mathrm{A}}\ln(\mathrm{Price}_{\mathrm{it}}^{\mathrm{A}}) + \varpi_{2\mathrm{t}}^{\mathrm{A}}\ln(\mathrm{Price}_{\mathrm{it}}^{\mathrm{B}}) + \beta_{1t}^{\mathrm{A}}\ln(\mathrm{Volume}_{\mathrm{it}}^{\mathrm{A}}) + \beta_{2t}^{\mathrm{A}}\ln(\mathrm{Volume}_{\mathrm{it}}^{\mathrm{B}}) + \beta_{3\mathrm{t}}^{\mathrm{A}}\mathrm{AvgRating}_{\mathrm{it}}^{\mathrm{A}} + \beta_{4\mathrm{t}}^{\mathrm{A}}\mathrm{AvgRating}_{\mathrm{it}}^{\mathrm{B}} + \psi_i + \mu_i^A + \varepsilon_{\mathrm{it}}^{\mathrm{A}}, \quad \text{(i)}$$

and

$$\ln(\mathrm{Rank}_{\mathrm{it}}^{\mathrm{B}}) = \omega_{0\mathrm{t}}^{\mathrm{B}} + \omega_{1\mathrm{t}}^{\mathrm{B}}\ln(\mathrm{Price}_{\mathrm{it}}^{\mathrm{B}}) + \varpi_{2\mathrm{t}}^{\mathrm{B}}\ln(\mathrm{Price}_{\mathrm{it}}^{\mathrm{A}}) + \beta_{1t}^{\mathrm{B}}\ln(\mathrm{Volume}_{\mathrm{it}}^{\mathrm{B}}) + \beta_{2t}^{\mathrm{B}}\ln(\mathrm{Volume}_{\mathrm{it}}^{\mathrm{A}}) + \beta_{3\mathrm{t}}^{\mathrm{B}}\mathrm{AvgRating}_{\mathrm{it}}^{\mathrm{B}} + \beta_{4\mathrm{t}}^{\mathrm{B}}\mathrm{AvgRating}_{\mathrm{it}}^{\mathrm{A}} + \psi_i + \mu_i^B + \varepsilon_{\mathrm{it}}^{\mathrm{B}}. \quad \text{(ii)}$$

By differencing the data across two sites and taking the first difference between these two equations, we can eliminate the potential confounding effects from the unobserved book fixed effect ψ_i. However, to eliminate the unobserved book-site fixed effects μ_i^A and μ_i^B, we need to use the first difference to take another level of difference across two periods. Therefore the DID model becomes

$$\Delta\left[\ln\left(\text{Rank}_i^A\right) - \ln\left(\text{Rank}_i^B\right)\right] = \eta_0 + \eta_1^A \Delta\ln\left(\text{Price}_i^A\right) + \eta_1^B \Delta\ln\left(\text{Price}_i^B\right) + \alpha_1^A \Delta\ln\left(\text{Volume}_i^A\right) + \alpha_1^B \Delta\ln\left(\text{Volume}_i^B\right) + \alpha_2^A \Delta\text{AvgRating}_i^A + \alpha_2^B \Delta\text{AvgRating}_i^B + \varepsilon_i, \quad (1)$$

where Δ denotes the variable difference from period t to period $t{+}1$ for a book i at site A (`Amazon.com`) or B (`BN.com`). Such a model tells us how the changes of independent variables cause the change of sales rank.

Model (1) is the model specification used in Chevalier and Mayzlin [2006] and is used as one benchmark model in our study. A second benchmark model is the benchmark Model (1) with an additional consumer rating measure; the dispersion or standard deviation of ratings (*StdRating*), defined as

$$\Delta\left[\ln\left(\text{Rank}_i^A\right) - \ln\left(\text{Rank}_i^B\right)\right] = \eta_0 + \eta_1^A \Delta\ln\left(\text{Price}_i^A\right) + \eta_1^B \Delta\ln\left(\text{Price}_i^B\right) + \alpha_1^A \Delta\ln\left(\text{Volume}_i^A\right) + \alpha_1^B \Delta\ln\left(\text{Volume}_i^B\right) + \alpha_2^A \Delta\text{AvgRating}_i^A + \alpha_2^B \Delta\text{AvgRating}_i^B + \alpha_3^A \Delta\text{StdRating}_i^A + \alpha_3^B \Delta\text{StdRating}_i^B + \varepsilon_i. \quad (2)$$

Upon these benchmark models, we gradually add other variables to test our hypotheses. When estimating the models, we monitor the possible multicollinearity issue for the variance inflation factor (VIF) to be well below the harmful level [Mason and Perreault 1991].

4.2.2. Sentiment Divergence Metrics vs. Nontextual Metrics. To validate our proposed divergence sentiment measures and show their initial values, we add each measure, *SentiDvg_KL* or *SentiDvg_JS*, into these two benchmark models separately. Specifically, corresponding to the benchmark Model (1), two new models are estimated.

$$\Delta\left[\ln\left(\text{Rank}_i^A\right) - \ln\left(\text{Rank}_i^B\right)\right] = \eta_0 + \eta_1^A \Delta\ln\left(\text{Price}_i^A\right) + \eta_1^B \Delta\ln\left(\text{Price}_i^B\right) + \alpha_1^A \Delta\ln\left(\text{Volume}_i^A\right) + \alpha_1^B \Delta\ln\left(\text{Volume}_i^B\right) + \alpha_2^A \Delta\text{AvgRating}_i^A + \alpha_2^B \Delta\text{AvgRating}_i^B + \beta_1^A \Delta \mathit{SentiDvg\_KL}_i^A + \beta_1^B \Delta \mathit{SentiDvg\_KL}_i^B + \varepsilon_i. \quad (1.1)$$

$$\Delta\left[\ln\left(\text{Rank}_i^A\right) - \ln\left(\text{Rank}_i^B\right)\right] = \eta_0 + \eta_1^A \Delta\ln\left(\text{Price}_i^A\right) + \eta_1^B \Delta\ln\left(\text{Price}_i^B\right) + \alpha_1^A \Delta\ln\left(\text{Volume}_i^A\right) + \alpha_1^B \Delta\ln\left(\text{Volume}_i^B\right) + \alpha_2^A \Delta\text{AvgRating}_i^A + \alpha_2^B \Delta\text{AvgRating}_i^B + \beta_1^A \Delta \mathit{SentiDvg\_JS}_i^A + \beta_1^B \Delta \mathit{SentiDvg\_JS}_i^B + \varepsilon_i. \quad (1.2)$$

Corresponding to the benchmark Model (2), the following two new models are estimated.

$$\Delta[\ln(\text{Rank}_i^A) - \ln(\text{Rank}_i^B)] = \eta_0 + \eta_1^A \Delta\ln(\text{Price}_i^A) + \eta_1^B \Delta\ln(\text{Price}_i^B) + \alpha_1^A \Delta\ln(\text{Volume}_i^A) + \alpha_1^B \Delta\ln(\text{Volume}_i^B) + \alpha_2^A \Delta\text{AvgRating}_i^A + \alpha_2^B \Delta\text{AvgRating}_i^B + \alpha_3^A \Delta\text{StdRating}_i^A + \alpha_3^B \Delta\text{StdRating}_i^B + \beta_1^A \Delta \mathit{SentiDvg\_KL}_i^A + \beta_1^B \Delta \mathit{SentiDvg\_KL}_i^B + \varepsilon_i, \quad (2.1)$$

$$\Delta[\ln(\text{Rank}_i^A) - \ln(\text{Rank}_i^B)] = \eta_0 + \eta_1^A \Delta\ln(\text{Price}_i^A) + \eta_1^B \Delta\ln(\text{Price}_i^B) + \alpha_1^A \Delta\ln(\text{Volume}_i^A) + \alpha_1^B \Delta\ln(\text{Volume}_i^B) + \alpha_2^A \Delta\text{AvgRating}_i^A + \alpha_2^B \Delta\text{AvgRating}_i^B + \alpha_3^A \Delta\text{StdRating}_i^A + \alpha_3^B \Delta\text{StdRating}_i^B + \beta_1^A \Delta \mathit{SentiDvg\_JS}_i^A + \beta_1^B \Delta \mathit{SentiDvg\_JS}_i^B + \varepsilon_i. \quad (2.2)$$

Table III presents the results of model estimations for both the periods between November 2009 and January 2010, and the period between November 2009 and July

Table III. Sentiment Divergence Metrics vs. Nontextual Review Metrics

	Nov 2009–Jan 2010						Nov 2009–Jul 2010					
	Mode 1(1)	Mode 1(2)	Model (1.1)	Model (1.2)	Model (2.1)	Model (2.2)	Model 1(1)	Model 1(2)	Model (1.1)	Model (1.2)	Model (2.1)	Model (2.2)
Amzn Δln(Price)	.009	−.006	.005	.006	−.015	−.017	.037	.037	.033	.037	.029	.032
BN Δln(Price)	.047	.038	.042	.043	.038	.037	−.007	−.009	−.010	−.008	−.012	−.009
Amzn Δln(Volume)	.107*	.080	.153**	.114*	.146**	.106	−.007	−.006	.018	.030	.020	.031
BN Δln(Volume)	−.083	−.077	−.049	−.058	−.015	−.016	−.023	.002	.071	.037	.131	.090
Amzn ΔAvgRating	−.042	−.189*	.021	−.007	−.069	−.090	−.013	−.033	.052	.043	.069	.071
BN ΔAvgRating	−.011	−.042	.010	−.013	−.059	−.107	.041	−.063	.078	.059	−.089	−.097
Amzn ΔStdRating		−.171*			−.098	−.092		−.022			.023	.036
BN ΔStdRating		−.030			−.104	−.133		−.136			−.230**	−.213**
Amzn ΔSentiDvg_KL			−.193***		−.193***				−.226***		−.251***	
BN ΔSentiDvg_KL			.052		.066				.118**		.142**	
Amzn ΔSentiDvg_JS				−.156***		−.153***				−.230***		−.253***
BN ΔSentiDvg_JS				.074		.113*				.065		.089
N	345	345	345	345	345	345	345	345	345	345	345	345
R-squared	.019	.030	.055	.049	.063	.058	.004	.011	.060	.054	.078	.070
Adjusted R-square	.001	.006	.033	.026	.035	.030	−.014	−.013	.037	.032	.050	.043
Model Fit *F*-stats	1.074	1.280	2.464**	2.165**	2.237**	2.051**	.227	.456	2.673***	2.415**	2.827***	2.529***

Note: The table lists the standardized coefficients of parameter estimates.
$^{*}p < .1$; $^{**}p < .05$; $^{***}p < .01$.

2010. As shown in the table, comparing with the benchmark Models (1) and (2), which only include nontextual review measures, the models including our divergence-based textual measures (Models (1.1)–(2.2)) have a significantly better fit. The model fit *F*-stats are not significant in most benchmark models but become significant in the models with the sentiment divergence measures.

More importantly, sentiment divergence metrics are highly significant. The dependent variables in the DID models are the relative sale ranks of books on `Amazon.com` (comparing to `BN.com`) over time. If the divergence sentiment metrics have positive sales effects, then we would expect a negative sign on the coefficients of divergence sentiment metrics, β_1^{A}, for reviews of `Amazon.com` and an opposite sign on the coefficients of divergence sentiment metrics, β_1^{B}, for reviews of `BN.com`. In other words, an increase in the sentiment divergence of reviews at Amazon over time leads to higher relative sales (lower relative sales rank) of books at Amazon over time. In contrast, an increase in the sentiment divergence of reviews at `BN.com` over time leads to higher relative sales (or smaller relative sales rank numbers) of books at `BN.com` over time, and thus lower relative sales (or bigger relative sales rank numbers) of books at `Amazon.com` over time.

In Table III, both metrics are significantly negative at `Amazon.com`. Specifically, an increase in the sentiment divergence of reviews at Amazon over time leads to higher relative sales (smaller relative sales rank numbers) of books at Amazon over time. For reviews at `BN.com`, the *SentiDvg_JS* measure is significantly positive in the period between November 2009 and January 2010 (Models (1.2) and (2.2)). The *SentiDvg_KL* measure at `BN.com` is also significantly positive in the period between November 2009 and July 2010 (Models (1.2) and (2.2)). This suggests that an increase in the sentiment divergence of reviews at `BN.com` over time leads to lower relative sales (or bigger relative sales rank numbers) of books at `Amazon.com` or higher relative sales of books at `BN.com` over time. Therefore, in both time spans we find a consistently positive effect of sentient divergence measures on sales at their own websites over time. This provides support for our hypotheses H1 and H2b. Overall, these results show that (1) textual content of consumer reviews does have additional positive impact on product sales in addition to the sales effect of nontextual review measures, and (2) our sentiment divergence measures are valid metrics to capture such impact.

Regarding the effects of consumer-rating variance, it is significantly negative at Amazon.com in model (2) in the period between Nov 2009 and Jan 2010, and it becomes insignificant in the period between Nov 2009 and July 2010. In contrast, it is insignificant at bn.com in the period between Nov 2009 and Jan 2010 and becomes significantly negative in models (2.1) and (2.2) in the period between Nov 2009 and July 2010. Thus, in an earlier period, an increase in the consumer rating variance at Amazon leads to higher relative sales change (or smaller relative sales rank number change) of books at Amazon. However, in a later period, an increase in the consumer-rating variance at bn.com leads to higher relative sales change (or smaller relative sales rank number change) of books at Amazon, or lower relative sales change of books at bn. This shows that the impact of consumer-rating variance on single-website sales shifts from positive to negative over time, which provides support for our hypothesis H2a. This result also further demonstrates the unique power of our sentiment divergence measures.

4.2.3. Sentiment Divergence Metrics vs. Alternative Sentiment Metrics. While we have shown that the sentiment divergence measures are helpful in predicting product sales, it is worthwhile investigating how this set of measures are different from existing sentiment measures in this task. As we have reviewed, subjectivity and polarity (valence) are two major types of measures in previous sentiment analysis literature, whose impacts on product sales have been studied [Ghose and Ipeirotis forthcoming; Liu et al. 2010].

Therefore, we compare these two alternative sentiment metrics with our proposed sentiment divergence measures.

Following previous research, we take an aggregation approach to assessing the subjectivity and polarity of product reviews. For review subjectivity, we apply OpinionFinder [Wilson et al. 2005] to classify subjectivity for each sentence in product reviews (subjective = 1 and objective = 0) and then average it to the product level, such that

$$Subj_i = \frac{1}{C_i}\sum_{j=1}^{C_i}\frac{\sum_{sentence_k \in v_i^j} Subjectivity_k}{\#\ sentences\ in\ v_i^j}.$$

To assess review polarity, we define the sentiment valence measure as the average word polarity score over all reviews for product i. Specifically,[5]

$$SentiValence_i = \frac{1}{C_i}\sum_{j=1}^{C_i}\frac{\sum_{w_k \in v_i^j} ms_k}{\#\, words\, in\, v_i^j}.$$

We use two different techniques to compute *ms* in the preceding formula, which give rise to two different versions of *SentiValence*: (1) *SentiValence_OF* in which we apply OpinionFinder to classify word-level sentiment to positive, negative, and neutral, and map them to *ms* values (1, −1, and 0, respectively). (2) *SentiValence_SW* in which we employ SentiWordNet in the same way as described in Section 3.1 to label every word with positive, negative, and neutral tags. These two methods show the use of learning-based and lexicon-based methods for sentiment assessment in an aggregation paradigm.

Correspondingly, three new models (Models (3), (4), and (5)) are constructed by adding into the benchmark Model (2) an additional sentiment measure. Specifically, Model (3) has *Subj*, Model (4) has *SentiValence_OF*, and Model (5) has *SentiValence_SW*.

$$\begin{aligned}\Delta[\ln(\mathrm{Rank}_i^A) - \ln(\mathrm{Rank}_i^B)] &= \eta_0 + \eta_1^A \Delta\ln(\mathrm{Price}_i^A) + \eta_1^B \Delta\ln(\mathrm{Price}_i^B)\\ &+ \alpha_1^A \Delta\ln(\mathrm{Volume}_i^A) + \alpha_1^B \Delta\ln(\mathrm{Volume}_i^B) + \alpha_2^A \Delta\mathrm{AvgRating}_i^A + \alpha_2^B \Delta\mathrm{AvgRating}_i^B\\ &+ \alpha_3^A \Delta\mathrm{StdRating}_i^A + \alpha_3^B \Delta\mathrm{StdRating}_i^B + \gamma_1^A \Delta\mathrm{Subj}_i^A + \gamma_1^B \Delta\mathrm{Subj}_i^B + \varepsilon_i \end{aligned} \tag{3}$$

$$\begin{aligned}\Delta[\ln(\mathrm{Rank}_i^A) - \ln(\mathrm{Rank}_i^B)] &= \eta_0 + \eta_1^A \Delta\ln(\mathrm{Price}_i^A) + \eta_1^B \Delta\ln(\mathrm{Price}_i^B)\\ &+ \alpha_1^A \Delta\ln(\mathrm{Volume}_i^A) + \alpha_1^B \Delta\ln(\mathrm{Volume}_i^B) + \alpha_2^A \Delta\mathrm{AvgRating}_i^A + \alpha_2^B \Delta\mathrm{AvgRating}_i^B\\ &+ \alpha_3^A \Delta\mathrm{StdRating}_i^A + \alpha_3^B \Delta\mathrm{StdRating}_i^B + \omega_1^A \Delta\mathrm{SentiValence}_i^A + \omega_1^B \Delta\mathrm{SentiValence}_i^B + \varepsilon_i \end{aligned} \tag{4)/(5}$$

To demonstrate the unique value of our sentiment divergence metrics relative to these three alternative metrics, we first show that when a single sentiment measure is deployed, sentiment divergence metrics are better than these three alternatives in capturing the sales effects of review textual content. Table IV presents the estimation results for Models (3), (4), and (5) with three alternative review textual metrics. In Model (3), none of the coefficients of the subjectivity metric is significant. The model fit statistics are not significant either. In Model (4) with the *SentiValence_OF* metric, the model fit statistics are not significant either. In Model (5), however, the coefficient of *SentiValence_SW* and the model fit statistic are significant in the period over November

[5]An aggregation approach is more suitable than a direct classification approach here, because the latter needs a large amount of training data, typically with rating as the label. Thus, a powerful polarity classification model will just mimic ratings, which will cause a multicollinearity problem in our DID model.

Table IV. Alternative Review Sentiment Metrics

	Nov 2009–Jan 2010			Nov 2009–Jul 2010		
	Model (3)	Model (4)	Model (5)	Model (3)	Model (4)	Model (5)
Amzn Δln(Price)	−.002	−.006	−.003	.033	.029	.051
BN Δln(Price)	.038	.031	.039	−.012	−.001	−.017
Amzn Δln(Volume)	.088	.084	.125*	.015	.019	.028
BN Δln(Volume)	−.099	−.083	−.038	.004	−.014	.025
Amzn ΔAvgRating	−.223**	−.239**	−.153	−.016	.069	.003
BN ΔAvgRating	−.025	−.037	−.048	−.049	−.049	−.064
Amzn ΔStdRating	−.204**	−.198**	−.184**	−.023	.031	−.028
BN ΔStdRating	.010	−.022	−.065	−.134	−.124	−.162*
Amzn ΔSubj	.033			.081		
BN ΔSubj	−.087			−.063		
Amzn ΔSentiValence		.074	−.160***		−.171***	−.14**
BN ΔSentiValence		.022	.061		−.023	.038
N	345	345	345	345	345	345
R-squared	.035	.034	.051	.019	.035	.027
Adjusted R-squared	.006	.005	.022	−.010	.006	−.002
Model Fit *F*-stats	1.214	1.183	1.791*	.664	1.210	.917

Note: The SentiValence measure is based on Opinion Finder (*SentiValence_OF*) in Model (4) and based on SentiWorldNet (*SentiValence_SW*) in Model (5).
Note: The table lists the standardized coefficients of parameter estimates.
*p < .1; **p < .05; ***p < .01.

2009 to January 2010. Thus, to demonstrate the superiority of our sentiment divergence metrics to *SentiValence_SW*, we conduct a nonnested *J*-test between Model (5) and Model (2.1) or Model (2.2). In a *J*-test, the predicted value of one model is included as an independent variable into another one to investigate whether it can bring further prediction power, as compared with existing variables. Table V presents the results of the *J*-test. For both time spans, the *J*-test rejects Model (5) as a better "true" model than either Model (2.1) or (2.2), (i.e., the predicted value of Model (5) is not significant when added into Model (2.1) or (2.2)) and accepts Model (2.1) or Model (2.2) as a better "true" model than Model (5). Therefore, these results show that our sentiment divergence metrics are more appropriate than the commonly used sentiment valence measures in capturing the sales effect of review content.[6]

At last, to further examine the effectiveness of the sentiment divergence measure, we inspect whether sentiment divergence metrics still have significant sales effects when added to a comprehensive model with all existing nontextual and textual sentiment measures. Models (6) and (7), corresponding to *SentiValence_OF* and *SentiValence_SW*, respectively, serve as the baseline.

$$\begin{aligned}\Delta[\ln(\mathrm{Rank}_i^A) - \ln(\mathrm{Rank}_i^B)] &= \eta_0 + \eta_1^A \Delta\ln(\mathrm{Price}_i^A) + \eta_1^B \Delta\ln(\mathrm{Price}_i^B)\\ &+ \alpha_1^A \Delta\ln(\mathrm{Volume}_i^A) + \alpha_1^B \Delta\ln(\mathrm{Volume}_i^B) + \alpha_2^A \Delta\mathrm{AvgRating}_i^A + \alpha_2^B \Delta\mathrm{AvgRating}_i^B\\ &+ \alpha_3^A \Delta\mathrm{StdRating}_i^A + \alpha_3^B \Delta\mathrm{StdRating}_i^B + \gamma_1^A \Delta\mathrm{Subj}_i^A + \gamma_1^B \Delta\mathrm{Subj}_i^B + \omega_1^A \Delta\mathrm{SentiValence}_i^A\\ &+ \omega_1^B \Delta\mathrm{SentiValence}_i^B + \varepsilon_i. \end{aligned} \tag{6)/(7}$$

[6]Our research focuses on explanation instead of prediction; therefore a low R^2 is not the concern. What matters is the significance of certain explanatory factors. For example, Chevalier and Mayzlin [2006] also have a low R^2 (less than 0.1) in their DID model. Another example is event studies in finance literature (e.g., [Asquith and Mullins 1986; Chaney et al. 1991; Holthausen and Leftwich 1986]), where most reported R^2 values are below 0.1. Their purposes, similar to ours, are to show the significant influence of certain factors on firm performance (stock price, in this case) in a noisy environment.

Table V. J-test: Sentiment Divergence Metrics vs. SentiWordNet-based Sentiment Valence

	H_0: The "true" model is							
	Nov 2009–Jan 2010				Nov 2009–Jul 2010			
	Model (5)	Model (5)	Model (2.1)	Model (2.2)	Model (5)	Model (5)	Model (2.1)	Model (2.2)
Amzn Δln(Price)	.006	.004	−.012	−.011	−.004	.000	.027	.028
BN Δln(Price)	−.003	.005	.024	.013	.000	−.001	−.012	−.008
Amzn Δln(Volume)	.003	.034	.100	.040	−.008	−.002	.019	.030
BN Δln(Volume)	−.011	−.006	.004	.017	.004	.005	.130*	.088
Amzn ΔAvgRating	−.026	−.038	−.023	−.002	−.017	−.009	.069	.072
BN ΔAvgRating	.024	.005	−.042	−.080	.031	.017	−.085	−.089
Amzn ΔStdRating	−.017	−.047	−.049	−.009	−.006	−.005	.024	.037
BN ΔStdRating	.034	.002	−.084	−.102	.024	.007	−.223*	−.195*
Amzn ΔSentiValence	.001	−.059			.021	−.002		
BN ΔSentiValence	.081	.073			.091	.062		
Amzn ΔSentiDvg_KL			−.150*				−.248***	
BN ΔSentiDvg_KL			.068				.142**	
Amzn ΔSentiDvg_JS				−.089				−.245***
BN ΔSentiDvg_JS				.112				.089
Predicted value from Model (2.1)	.265**				.317***			
Predicted value from Model (2.2)		.205**				.277***		
Predicted value from Model (5)			.082	.140			.008	.019
Predicted value from Model (5)			.082	.140			.008	.019
N	345	345	345	345	345	345	345	345
R-squared	.068	.063	.064	.063	.086	.074	078	.071
Adjusted R-squared	.037	.032	.033	.032	.055	.043	.048	.040
Model Fit *F*-stats	2.194**	2.048**	2.076**	2.022**	2.834***	2.414***	2.564***	2.298**

Note: The table lists the standardized coefficients of parameter estimates.
*p < .1; **p <. 05; ***p < .01.

In comparison, Models (6.1), (6.2), (7.1), and (7.2) add the two sentiment divergence measures separately.

$$\begin{aligned}\Delta[\ln(\mathrm{Rank}_i^A) - \ln(\mathrm{Rank}_i^B)] &= \eta_0 + \eta_1^A \Delta\ln(\mathrm{Price}_i^A) + \eta_1^B \Delta\ln(\mathrm{Price}_i^B)\\ &+ \alpha_1^A \Delta\ln(\mathrm{Volume}_i^A) + \alpha_1^B \Delta\ln(\mathrm{Volume}_i^B) + \alpha_2^A \Delta\mathrm{AvgRating}_i^A + \alpha_2^B \Delta\mathrm{AvgRating}_i^B\\ &+ \alpha_3^A \Delta\mathrm{StdRating}_i^A + \alpha_3^B \Delta\mathrm{StdRating}_i^B + \gamma_1^A \Delta\mathrm{Subj}_i^A + \gamma_1^B \Delta\mathrm{Subj}_i^B + \omega_1^A \Delta\mathrm{SentiValence}_i^A\\ &+ \omega_1^B \Delta\mathrm{SentiValence}_i^B + \beta_1^A \Delta \mathit{SentiDvg\_KL}_i^A + \beta_1^B \Delta \mathit{SentiDvg\_KL}_i^B + \varepsilon_i, \qquad (6.1)/(7.1)\end{aligned}$$

$$\begin{aligned}\Delta[\ln(\mathrm{Rank}_i^A) - \ln(\mathrm{Rank}_i^B)] &= \eta_0 + \eta_1^A \Delta\ln(\mathrm{Price}_i^A) + \eta_1^B \Delta\ln(\mathrm{Price}_i^B)\\ &+ \alpha_1^A \Delta\ln(\mathrm{Volume}_i^A) + \alpha_1^B \Delta\ln(\mathrm{Volume}_i^B) + \alpha_2^A \Delta\mathrm{AvgRating}_i^A + \alpha_2^B \Delta\mathrm{AvgRating}_i^B\\ &+ \alpha_3^A \Delta\mathrm{StdRating}_i^A + \alpha_3^B \Delta\mathrm{StdRating}_i^B + \gamma_1^A \Delta\mathrm{Subj}_i^A + \gamma_1^B \Delta\mathrm{Subj}_i^B + \omega_1^A \Delta\mathrm{SentiValence}_i^A\\ &+ \omega_1^B \Delta\mathrm{SentiValence}_i^B + \beta_1^A \Delta \mathit{SentiDvg\_JS}_i^A + \beta_1^B \Delta \mathit{SentiDvg\_JS}_i^B + \varepsilon_i. \qquad (6.2)/(7.2)\end{aligned}$$

Table VI presents the results of this analysis. First, we find that compared to Models (6) and (7), models with additional sentiment divergence metrics have significantly higher model fit (higher R-squared, adjusted R-squared, and F-stats). Second, and more importantly, the coefficients are also more significant for our sentiment divergence metrics than for competing metrics in these models. Finally, the standardized coefficients in Table VI show us the relative effect of each independent variable on the dependent variable [Hair et al. 1995]. In Models (6.1) (7.2), the magnitudes of the

Table VI. Overall Comparison

	Nov 2009–Jan 2010						Nov 2009–Jul 2010					
	Model (6)	Model (7)	Model (6.1)	Model (6.2)	Model (7.1)	Model (7.2)	Model (6)	Model (7)	Model (6.1)	Model (6.2)	Model (7.1)	Model (7.2)
Amzn Δln(Price)	−.001	−.001	−.011	−.010	−.017	−.016	.024	.046	.034	.030	.022	.019
BN Δln(Price)	.030	.039	.036	.037	.027	.028	−.004	−.023	−.014	−.016	−.006	−.008
Amzn Δln(Volume)	.094	.137**	.125*	.150**	.111	.142**	.044	.059	.049	.020	.064	.047
BN Δln(Volume)	−.110	−.059	−.023	−.022	−.029	−.016	−.013	.031	.097	.152**	.073	.124*
Amzn ΔAvgRating	−.283**	−.166	−.113	−.103	−.148	−.124	.104	.032	.082	.074	.163*	.170*
BN ΔAvgRating	−.024	−.034	−.098	−.040	−.151	−.098	−.039	−.050	−.077	−.070	−.082	−.087
Amzn ΔStdRating	−.240	−.199**	−.132	−.124	−.129	−.129	.040	−.029	.032	.027	.075	.071
BN ΔStdRating	.025	−.037	−.122	−.084	−.139	−.114	−.126	−.163*	−.216**	−.243***	−.201**	−.228**
Amzn ΔSubj	.076	.051	−.085	.011	.064	.004	−.191***	0.100*	−.023	.058	−.138**	.085
BN ΔSubj	.039	−.066	.084	−.001	.093	−.013	−.009	−.081	.060	.003	.028	−.036
Amzn ΔSentiValence	.033	−.159***	.026	−.014	.008	.065	.112**	−.161***	.073	.007	.094*	−.145**
BN ΔSentiValence	−.099	.061	−.012	.102	−.026	.120	−.061	.029	−.037	.097	−.053	.065
Amzn ΔSentiDvg_KL				−.179*		−.170***				−.273***		−.231***
BN ΔSentiDvg_KL				.108		.153				.163**		.154**
Amzn ΔSentiDvg_JS			−.093		−.135**				−.249***		−.223***	
BN ΔSentiDvg_JS			.121		.165*				.084		.085	
N	345	345	345	345	345	345	345	345	345	345	345	345
R-squared	.041	.056	.065	.069	.066	.072	.048	.040	.079	.089	.092	.102
Adjusted R-squared	.006	.021	.025	.029	.027	.032	.014	.005	.040	.050	.053	.064
Model Fit *F*-stats	1.177	1.626*	1.636*	1.743**	1.679*	1.822**	1.401	1.149	2.030**	2.297***	2.375***	2.676***

Note: The SentiValence measure is based on Opinion Finder (*SentiValence_OF*) in Model (6) and on SentiWorldNet (*SentiValence_SW*) in Model (7).
Note: The table lists the standardized coefficients of parameter estimates.
$^{*}p < .1$; $^{**}p < .05$; $^{***}p < .01$.

standardized coefficients of sentiment divergence metrics are much bigger than those of other sentiment metrics, that is, the sentiment divergence metrics have much bigger relative impacts on product sales than the other two types of textual metrics. Overall, these results further demonstrate the power of our sentiment divergence measures in capturing the sales effects of textual review content.

5. CONCLUSIONS, IMPLICATIONS, AND FUTURE RESEARCH

By integrating marketing theories with text-mining techniques, this study tackles an intriguing and challenging business intelligence problem. Specifically, upon proposing a set of innovative text-based measures, we first found a significant effect of WOM's sentiment divergence on product sales performance, based on a large empirical data set. Second, this effect is not fully captured by traditional nontextual review measures. Furthermore, our divergence metrics are shown to better capture the sales effect of review content and, therefore, are superior to commonly used textual measures in the literature.

These findings have important practical and theoretical implications. With the rapid expansion of social media, their influence on firm strategy and product sales has received a growing amount of attention from researchers and managers. Our results provide important insights for firms managing social media. The strong effect of sentiment divergence on product sales shows that firms can develop appropriate strategies for managing the sentiment divergence in review content. An important debate that has arisen recently in practice is whether e-commerce firms should allow consumers to post uncensored reviews of their products and services, and how much information they should allow consumers to post [Woods 2002]. Due to the concern that negative reviews might hurt sales, many analysts and practitioners suggest that sellers, if they decide to allow consumers to post reviews, should adopt a survey model allowing consumers to post only ratings instead of detailed content [Woods 2002]. However, our results show that such a model may not be valid. The main value of consumer reviews may lie in their textual content. More importantly, negative textual content might not be necessarily bad for the sellers. A negative review may still provide important matching information for other consumers and increase their purchase intention. What really matters is the divergence or dispersion of the review content. More divergent opinions exhibited in the textual content can lead to higher product sales, even though some of the reviews are negative. Most recently, Facebook updated its "Like" button function by giving individuals an option to make comments and switched from a pure rating evaluation system to a hybrid system with both ratings and textual comments [Lavrusik 2011]. This seems to further validate the importance of the textual review content. For online retailers or e-commerce firms, they can use different strategies to increase the potential sentiment divergence in reviews. One possibility, as CNET did, is to provide separate "Pros" and "Cons" sections in addition to the overall comments section, to solicit divergent opinions from consumers.

Our results show that to capture the sales effects of review textual content, it is necessary to capture the sentiment dispersion/divergence aspect, instead of solely the central tendency or polarity aspects studied in the literature. Such divergence measures may provide useful insights about the impact of textual information in social media in general and about the emerging research on consumer reviews or user-generated content, in particular. Our focus in the current article is to propose some valid text-based measures to capture the sales effect of WOM information in social media. Once we demonstrate such validity, another application is to use the sentiment divergence metrics to predict product sales. It is important to note that at the computational level, the current set of text-based measures are very simple and parsimonious, in the sense that they only exploit shallow lexical information in the review text and transform such into approximated sentiment distributions. Therefore, this is a very conservative way to demonstrate the sales effects of review sentiment divergence. Given the significant sales effects of these parsimonious and even naïve metrics, we would expect a stronger sales effect from some more comprehensive metrics incorporating nuanced information hidden in text (e.g., product feature-level sentiment diversity, relative strength or relevance of discussions). For sales prediction purposes, future research could explore whether more sophisticated natural language-processing techniques would be helpful in better capturing this nuanced information to make a better prediction.

Some recent research suggests the existence of self-selection biases in online product reviews (i.e., the early adopters' opinions may differ from those of the broader consumer population) and shows that consumers do not fully account for such biases when making purchase decisions [Li and Hitt 2008]. In light of such findings, opportunities exist for future research to strengthen our model by computing all variables on a subset of reviews (the most recent, the oldest, or the most "helpful" deemed by readers), instead of all reviews associated with a given product.

When empirically testing the validity of our sentiment divergence measure, in alignment with previous literature, we chose a setting in which both nontextual consumer ratings and textual content information were available. However, in reality, quite often only textual content is available for WOM in social media (e.g., message boards, chat rooms, and blogs). Future research can study how our textual metrics would influence product sales or other financial performances in those settings. We suspect the impact of these measures might be even stronger there since nontextual, discrete rating information can lead to consumer herding behavior by ignoring detailed information from review content [Bikhchandani et al. 1992].

REFERENCES

ABBASI, A., CHEN, H. C., AND SALEM, A. 2008. Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums. *ACM Trans. Info. Syst. 26*, 3, 1–34.

AIROLDI, E., BAI, X., AND PADMAN, R. 2006. Markov blankets and meta-heuristics search: Sentiment extraction from unstructured texts. Lecture Notes in Artificial Intelligence, vol. 3932, 167–187.

ANDREAS, K. M. AND MICHAEL, H. 2010. Users of the world, unite! The challenges and opportunities of social media. *Bus. Horizons 53*, 1, 59–68.

ANTWEILER, W. AND FRANK, M. Z. 2004. Is all that talk just noise? The information content of internet stock message boards. *J. Finance 59*, 3, 1259–1294.

ASQUITH, P. AND MULLINS, D. 1986. Equity issues and offering dilution. *J. Finan. Econ. 15*, 1–2, 61–89.

BAGNOLI, M., BENEISH, M. D., AND WATTS, S. G. 1999. Whisper forecasts of quarterly earnings per share. *J. Account. Econ. 28*, 1, 27–50.

BIKHCHANDANI, S., HIRSHLEIFER, D., AND WELCH, I. 1992. A theory of fads, fashion, custom, and cultural-change as informational cascades. *J. Polit. Econ. 100*, 5, 992–1026.

BOIY, E. AND MOENS, M. F. 2009. A machine learning approach to sentiment analysis in multilingual web texts. *Infor. Retriev. 12*, 5, 526–558.

CHANEY, P. K., DEVINNEY, T. M., AND WINER, R. S. 1991. The impact of new product introductions on the market value of firms. *J. Business 64*, 4, 573–610.

CHEN, H. 2010. Business and market intelligence 2.0. *IEEE Intell. Syst. 25*, 6, 2–5.

CHEN, Y. AND XIE, J. 2008. Online consumer review: Word-of-mouth as a news element of marketing communication mix. *Manag. Sci. 54*, 3, 477–491.

CHEN, Y., WANG, Q., AND XIE, J. 2011. Online social interactions: A natural experiment on word-of-mouth versus observational learning. *J. Market. Res.* To appear.

CHEVALIER, J. AND GOOLSBEE, A. 2003. Measuring prices and price competition online: Amazon.com and BarnesandNoble.com. *Quant. Market. Econ. 1*, 203–222.

CHEVALIER, J. A. AND MAYZLIN, A. 2006. The effect of word-of-mouth on sales: Online book reviews. *J. Market. Res. 43*, 3, 345–354.

CHUNG, W. 2009. Automatic summarization of customer reviews: An integrated approach. In *Proceedings of the Americas Conference on Information Systems*.

DAS, S., MARTINEZ-JEREZ, A., AND TUFANO, P. 2005. eInformation: A clinical study of investor discussion and sentiment. *Financ. Manag. 34*, 3, 103–137.

DAS, S. R. AND CHEN, M. Y. 2007. Yahoo! for Amazon: Sentiment extraction from small talk on the web. *Manag. Science 53*, 9, 1375–1388.

DAVE, K., LAWRENCE, S., AND PENNOCK, D. M. 2003. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceedings of the International Conference on the World Wide Web*.

DELLAROCAS, C. 2003. The digitization of word-of-mouth: Promise and challenges of online feedback mechanisms. *Manag. Science 49*, 10, 1407–1424.

DELLAROCAS, C., ZHANG, X. Q., AND AWAD, N. F. 2007. Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *J. Interact. Market. 21*, 4, 23–45.

DUAN, W. J., GU, G., AND WHINSTON, A. B. 2008. Do online reviews matter? An empirical investigation of panel data. *Decision Support Syst. 45*, 4, 1007–1016.

EFRON, M. 2004. Cultural orientation: Classifying subjective documents by vocation analysis. In *Proceedings of the AAAI Fall Symposium on Style and Meaning in Language, Art, and Music*.

ESULI, A. AND SEBASTIANI, F. 2006. SentiWordNet: A publicly available lexical resource for opinion mining. In *Proceedings of the 5th Conference on Language Resources and Evaluation*.

FORMAN, C., GHOSE, A., AND WIESENFELD, B. 2008. Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Info. Syst. Res. 19*, 3, 291–313.

GENTZKOW, M. AND SHAPIRO, J. M. 2006, Media bias and reputation. *J. Political Econ. 114*, 2, 280–316.

GAO, Y., MAO, C. X., AND ZHONG, R. 2006. Divergence of opinion and long-term performance of initial public offerings. *J. Financ. Res. 29*, 1, 113–129.

GHOSE, A. AND IPEIROTIS, P. G. Forthcoming. Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Trans. Knowl. Data Eng.*

GODES, D. AND MAYZLIN, D. 2004. Using online conversations to study word-of-mouth communication. *Market. Science 23*, 4, 545–560.

GODES, D., MAYZLIN, D., CHEN, Y., DAS, S., DELLAROCAS, C., PFEIFFER, B., LIBAI, B., SEN, S., SHI, M. Z., AND VERLEGH, P. 2005. The firm's management of social interactions. *Market. Letters 16*, 3-4, 415–428.

HAIR, J., ANDERSON, R., TATHAM, R., AND BLACK, W. 1995. *Multivariate Data Analysis*, Prentice-Hall.

HERR, P. M., KARDES, F. R., AND KIM, J. 1991. Effects of word-of-mouth and product-attribute information on persuasion—An accessibility-diagnosticity perspective. *J. Consum. Res. 17*, 4, 454–462.

HOLTHAUSEN, R. AND LEFTWICH, R. 1986. The effect of bond rating changes on common stock prices. *J. Financ. Econo. 17*, 1, 57–89.

HOTELLING, H. 1929. Stability in competition. *Econ. J. 39*, 153, 41–57.

HU, Y. AND LI, W. J. 2011. Document sentiment classification by exploring description model of topical terms. *Comput. Speech Lang. 25*, 2, 386–403.

KAHN, B. AND MEYER, R. 1991. Consumer multiattribute judgments under attribute uncertainty. *J. Consum. Res. 17*, 4, 508–522.

KAHNEMAN, D. AND TVERSKY, A. 1979. Prospect theory: An analysis of decision under risk. *Econometrica 47*, 3, 263–291.

KENNEDY, A. AND INKPEN, D. 2006. Sentiment classification of movie reviews using contextual valence shifters. *Comput. Intell. 22*, 2, 110–125.

KULLBACK, S. AND LEIBLER, R. A. 1951. On information and sufficiency. *Annals Math. Stat. 22*, 1, 79–86.

LAVRUSIK, V. 2011. Facebook like button takes over share button functionality. http://mashable.com/2011/02/27/facebook-like-button-takes-over-share-button-functionality/.

LI, N. AND WU, D. D. 2010. Using text mining and sentiment analysis for online forums hotspot detection and forecast. *Decision Support Syst. 48*, 2, 354–368.

LI, X. X. AND HITT, L. M. 2008. Self-selection and information role of online product reviews. *Info. Syst. Res. 19*, 4, 456–474.

LIN, J. H. 1991. Divergence measures based on the shannon entropy. *IEEE Trans. Info. Theory 37*, 1, 145–151.

LIU, B., HU, M., AND CHENG, J. 2005. Opinion observer: Analyzing and comparing opinions on the web. In *Proceedings of the 14th International Conference on the World Wide Web*. 342–351.

LIU, Y. 2006. Word-of-mouth for movies: Its dynamics and impact on box office revenue. *J. Market. 70*, 3, 74–89.

LIU, Y., CHEN, Y., LUSCH, R., CHEN, H., ZIMBRA, D., AND ZENG, S. 2010. User-generated content on social media: Predicting new product market success from online word-of-mouth. *IEEE Intell. Syst. 25*, 6, 8–12.

LIU, Y., HUANG, X., AN, A., AND YU, X. 2007. ARSA: A sentiment-aware model for predicting sales performance using blogs. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. 607–614.

MASON, C. H. AND PERREAULT, W. D. 1991. Collinearity, power, and interpretation of multiple-regression analysis. *J. Market. Res. 28*, 3, 268–280.

MISHNE, G. 2005. Experiments with mood classification in blog posts. In *Proceedings of the 1st Workshop on Stylistic Analysis of Text for Information Access*.

MISHNE, G. 2006. Multiple ranking strategies for opinion retrieval in blogs. In *Proceedings of the Text Retrieval Conference*.

MIZERSKI, R. W. 1982. An attribution explanation of the disproportionate influence of unfavorable information. *J. Consum. Res. 9*, 3, 301–310.

MULLAINATHAN, S. AND SHLEIFER, A. 2005. The market for news. *Amer. Econ. Rev. 95*, 1031–1053.

NASUKAWA, T. AND YI, J. 2003. Sentiment analysis: Capturing favorability using natural language processing. In *Proceedings of the International Conference on Knowledge Capture*. 70–77.

NIGAM, K. AND HURST, M. 2004. Towards a robust metric of opinion. In *Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text*. 598–603.

PANG, B. AND LEE, L. 2005. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proceedings of the Meeting of the Association for Computer Learning*. 115–124.

PANG, B. AND LEE, L. 2008. Opinion mining and sentiment analysis. *Foundat. Trends Info. Retriev. 2*, 1–2, 1–135.

PANG, B., LEE, L., AND VAITHYANATHAN, S. 2002. Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. 79–86.

POPESCU, A. M. AND ETZIONI, O. 2005. Extracting product features and opinions from reviews. In *Proceedings of the Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*. 339–346.

SARVARY, M. AND PARKER, P. 1997. Marketing information: A competitive analysis. *Market. Science 16*, 1, 24–38.

SUBASIC, P. AND HUETTNER, A. 2001. Affect analysis of text using fuzzy semantic typing. *IEEE Trans. Fuzzy Syst. 9*, 4, 483–496.

SUBRAHMANIAN, V. S. AND REFORGIATO, D. 2008. AVA: Adjective-verb-adverb combinations for sentiment analysis. *IEEE Intell. Syst. 23*, 4, 43–50.

SUN, M. 2009. How does variance of product ratings matter? (http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1400173).

SUROWIECKI, J. 2005. *The Wisdom of Crowds*. Anchor Books, New York.

TAN, S. B. AND ZHANG, J. 2008. An empirical study of sentiment analysis for chinese documents. *Expert Syst. Appl. 34*, 4, 2622–2629.

THELWALL, M., BUCKLEY, K., PALTOGLOU, G., CAI, D., AND KAPPAS, A. 2010. Sentiment strength detection in short informal text. *J. Amer. Soc. Info. Sci. Tec. 61*, 12, 2544–2558.

THOMAS, M., PANG, B., AND LEE, L. 2006. Get out the vote: Determining support or opposition from congressional floor-debate transcripts. In *Proceedings of the Conference on Empirical Methods on Natural Language Processing*. 327–335.

TUMARKIN, R. AND WHITELAW, R. F. 2001. News or noise? Internet postings and stock prices. *Financ. Anal. J. 57*, 3, 41–51.

TURNEY, P. D. 2001. Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*. 417–424.

WEST, P. AND BRONIARCZYK, S. 1998. Integrating multiple opinions: The role of aspiration level on consumer response to critic consensus. *J. Consum. Res. 25*, 1, 38–51.

WIEBE, J., WILSON, T., BRUCE, T., BELL, M., AND MARTIN, M. 2004. Learning subjective language. *Computat. Linguist. 30*, 3, 277–308.

WILSON, T., HOFFMANN, P., SOMASUNDARAN, S., KESSLER, J., WIEBE, J., CHOI, Y., CARDIE, C., RILOFF, E., AND PATWARDHAN, S. 2005. OpinionFinder: A system for subjectivity analysis. In *Proceedings of the Human Language Technology Conference / Conference on Empirical Methods in Natural Language Processing.*

WILSON, T., WIEBE, J., AND HOFFMANN, P. 2009. Recognizing contextual polarity: an exploration of features for phrase-level sentiment analysis. *Computat. Linguist. 35*, 3, 399–433.

WILSON, T., WIEBE, J., AND HWA, R. 2004. Just how mad are you? Finding strong and weak opinion clauses. In *Proceedings of the 19th National Conference on Artificial Intelligence.* 761–769.

WOODS, B. 2002. Should online shoppers have their say? *E-Commerce Times*. http://www.ecommercetimes.com/perl/story/20049.html.

WOOLDRIDGE, J. 2002. *Econometric Analysis of Cross Section and Panel Data*, MIT Press.

YANG, C., TANG, X., WONG, Y. C., AND WEI, C.-P. 2010. Understanding online consumer review opinion with sentiment analysis using machine learning. *Pacif. Asia J. Assoc. Info. Syst. 2*, 3, 7.

YI, J., NASUKAWA, T., BUNESCU, R., NIBLACK, W., AND CENTER, C. 2003. Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. In *Proceedings of the 3rd IEEE International Conference on Data Mining*. 427–434.

YU, H. AND HATZIVASSILOGLOU, V. 2003. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. 129–136.

ZHANG, C. L., ZENG, D., LI, J. X., WANG, F. Y., AND ZUO, W. L. 2009. Sentiment analysis of chinese documents: From sentence to document level. *J. Amer. Soc. Info. Sci. Tec. 60*, 12, 2474–2487.

ZHANG, Z. 2008. Weighing stars: Aggregating Online product reviews for intelligent e-commerce applications. *IEEE Intell. Syst. 23*, 5, 42–49.

ZHU, F. AND ZHANG, Z. 2010. Impacts of online consumer reviews on sales: The moderating role of product and consumer characteristics. *J. Market. 74*, 133–148.

Received January 2012; accepted January 2012