WENDY W. MOE and MICHAEL TRUSOV*

Research has shown that consumer online product ratings reflect both the customers' experience with the product and the influence of others' ratings. In this article, the authors measure the impact of social dynamics in the ratings environment on both subsequent rating behavior and product sales. First, they model the arrival of product ratings and separate the effects of social influences from the underlying (or baseline) ratings behavior. Second, the authors model product sales as a function of posted product ratings while decomposing ratings into a baseline rating, the contribution of social influence, and idiosyncratic error. This enables them to quantify the sales impact of observed social dynamics. The authors consider both the direct effects on sales and the indirect effects that result from the influence of dynamics on future ratings (and thus future sales). The results show that although ratings behavior is significantly influenced by previously posted ratings and can directly improve sales, the effects are relatively short lived once indirect effects are considered.

Keywords: online word of mouth, ratings, reviews, social dynamics, hazard models, Internet marketing

The Value of Social Dynamics in Online Product Ratings Forums

In recent years, online product ratings and reviews have taken on a larger role in the consumer decision process. Not only are more consumers contributing their opinions, but potential buyers are also increasingly relying on the information provided by others in these forums. The result is that online customer ratings have the potential to significantly affect product sales (Chevalier and Mayzlin 2006; Clemons, Gao, and Hitt 2006; Dellarocas, Zhang, and Awad 2007). In theory, online review forums facilitate the exchange of information and help consumers make more informed decisions. For example, Chen and Xie (2008) suggest that online reviews created by users can work as "sales assistants" to help novice consumers identify the products that best match their idiosyncratic preferences. The authors argue that, in the absence of review information, novice consumers may be less likely to buy a product if only seller-created product attribute information is available, suggesting that the availability of consumer-generated reviews may lead to an increase in sales.

The common underlying assumption of studies that investigate the impact of consumer reviews on product sales is that posted product ratings reflect the customers' experience with the product independent from the ratings of others. However, researchers have shown that posted product ratings are subject to a number of influences unrelated to a consumer's objective assessment of the product. For example, Schlosser (2005) shows that posted product ratings are influenced by social dynamics. Specifically, the rating a person posts for a product is affected by previously posted ratings. In addition, Godes and Silva (2009) demonstrate ratings dynamics that result in a negative trend in posted product ratings as the volume of postings increase. Li and Hitt (2008) observe a similar trend.

The consequence of these ratings dynamics is that user-provided product ratings do not always accurately reflect product performance, yet they still have the potential to significantly influence product sales. This can be disconcerting for product marketers, and as a result, many marketers are investing in activities intended to create a more favorable

*Wendy W. Moe is Associate Professor of Marketing, Robert H. Smith School of Business, University of Maryland (e-mail: wmoe@rhsmith.umd.edu). Michael Trusov is Assistant Professor of Marketing, Robert H. Smith School of Business, University of Maryland (e-mail: mtrusov@rhsmith.umd.edu). The authors thank the Marketing Science Institute for the research grant provided to this project. They also thank BazaarVoice for the data analyzed in this article as well as the invaluable conversations surrounding this topic. Christophe Van den Bulte served as associate editor for this article.

ratings environment for their products with the intention of boosting sales (Dellarocas 2006).

Our objective in this article is to measure the value of the social dynamics found in online customer rating environments. We explicitly model the arrival of posted product ratings and separate the effects of social dynamics on ratings from the underlying baseline ratings behavior (which we argue reflects the consumers' "socially unbiased" product evaluations).[1] In many ratings forums, consumers evaluate products along a five-star scale. Therefore, we capture the ratings process by modeling the arrival of ratings within each star level as five separate (but interrelated) hazard processes. This process enables us to capture the timing and the valence of posted ratings simultaneously. In addition, we include time-varying hazard covariates to capture the effect of social influence on ratings behavior. The resulting model estimates enable us to compute a set of ratings metrics that represent the expected ratings behavior both with and without the effect of social dynamics. We then decompose observed ratings into a baseline ratings component, a component that represents the impact of social dynamics, and an idiosyncratic error component. To capture the sales impact of social dynamics, we model sales as a function of these component-ratings metrics.

Our model results indicate that there are substantial social dynamics in the ratings environment. We study these dynamics further by examining both their direct and indirect effects on product sales. We show specifically that the dynamics observed in a product's ratings valence (or average rating) can have direct and immediate effects on sales. We also show that these dynamics can have additional indirect effects on future sales through their influence on future ratings. These indirect effects can mitigate the long-term impact of ratings dynamics on sales, particularly in cases in which the level of opinion variance is increased.

One unique aspect of our research that differentiates it from previous studies is that we model variation in sales over a large set of products over time. Many of the published studies in this area of research use publicly available data on ratings and product sales rank at a fixed point in time, allowing only for cross-sectional analyses (see, e.g., Chevalier and Mayzlin 2006). In contrast, we use a data set obtained from an online retailer that contains longitudinal ratings and sales data, which enables us to examine changes in product sales from period to period as a function of changes in the ratings environment. The longitudinal nature of the data also provides multiple observations for each product in the sample. This added richness in the data enables us to explicitly model product heterogeneity.

Several other aspects of our data are worth mentioning. First, the functionality to post ratings on the site was introduced approximately halfway through our data period. The weekly sales levels observed before the introduction of the ratings functionality enable us to estimate a product's baseline sales level absent any ratings effects. The second aspect is the product category. Previous researchers have typically focused on movies and books, likely because of the availability of data in these categories. However, movies and books are unique in that they have product life cycles that are both short and follow predictable exponential patterns (Moe and Fader 2001; Sawhney and Eliashberg 1996). These products experience the greatest level of sales (and ratings activity) immediately after launch, quickly after which sales (and ratings activity) taper off dramatically. The danger of using such product categories is that results can be sensitive to when in the product life cycle the researcher collects the data. In this article, we use sales and ratings data for products in a mature product category with relatively stable sales. As a result, the sales changes observed can be attributed to changes in the ratings environment and are less likely caused by the natural progression of the product life cycle or evolving consumer base (e.g., Li and Hitt 2008).

In the next section, we review the existing literature pertaining to the effects of online user-generated ratings. Included in this review is a discussion of previous research that has revealed how social influences and dynamics can affect the posting of product ratings. After the literature review, we present a conceptual framework that relates product sales to online consumer ratings. We then describe the data used in this article. In particular, we highlight some of the characteristics of the product category featured in our data that make it well suited for the research questions we address. We then develop the model and propose a set of metrics based on the ratings component of the model, which enables us to decompose the ratings effect and measure the sales impact of social dynamics. We then present the results and highlight some of the implications.

LITERATURE REVIEW

There is a growing body of research in both the marketing literature and the information technology literature that examines the effects of online word of mouth. The authors of these articles have considered various forms of online word of mouth, including user-provided ratings and reviews and newsgroup postings. Although some researchers have focused on measuring the effects of online word of mouth on performance measures (e.g., sales, growth, television viewership), others have studied the dynamics observed in these online word-of-mouth environments.

Effects of Online Word of Mouth

The majority of research in this area has identified three metrics of online word of mouth: valence, variance, and volume (Dellarocas and Narayan 2006). Valence is represented most frequently by an average rating measure (Chevalier and Mayzlin 2006; Clemons, Gao, and Hitt 2006; Dellarocas, Zhang, and Awad 2007; Duan, Gu, and Whinston 2008). It has also been represented by some measure of positivity (or negativity) of ratings (Chevalier and Mayzlin 2006; Godes and Mayzlin 2004; Liu 2006). The variance in ratings has also been measured in a variety of ways, ranging from a statistical variance (Clemons, Gao, and Hitt 2006) to entropy (Godes and Mayzlin 2004), and volume is represented most commonly by the number of postings.

Although several authors have focused on studying the effects of valence, variance, and volume of online word of mouth on product performance, they have found varying empirical results (see Table 1). Two articles (Duan, Gu, and Whinston 2008; Liu 2006) examine the impact of posted user reviews on movie box office sales. Although the effects

[1]In the subsequent presentation, we use the term "unbiased" in a narrow sense, specific to the discussed context.

Table 1
LITERATURE REVIEW

Article	*Product Category*	*Dependent Variable*	*Significant Word-of-Mouth Effects*
Liu (2006)	Movies	Sales	Number of posts
Duan, Gu, and Whinston (2008)	Movies	Sales	Number of posts
Dellarocas, Zhang, and Awad (2007)	Movies	Sales diffusion parameters	Average rating, number of ratings
Clemons, Gao, and Hitt (2006)	Beer	Sales growth rate	Average rating, standard deviation of ratings
Godes and Mayzlin (2004)	Television shows	Television-viewership ratings	Entropy of posts, number of posts
Chevalier and Mayzlin (2006)	Books	Sales rank	Average rating, number of ratings
Current study	Bath, fragrance, and beauty products	Cross-product temporal variation in ratings and sales	Static and dynamic effects of ratings

of valence and volume were modeled in both cases, only the volume of word of mouth was significant. Dellarocas, Zhang, and Awad (2007) also study the effects of online word of mouth on movie box office sales. However, rather than modeling weekly sales, they examine the patterns of sales growth and find that both valence and volume of online word of mouth have significant effects. Another study by Clemons, Gao, and Hitt (2006) focuses attention on the craft beer category, finding that the valence and variance (but not the volume) of ratings affect sales growth. In particular, the authors find that the valence of the top quartile of ratings has the greatest effect on predicting sales growth.

An important challenge that must be addressed when studying the sales effect of ratings is the potentially endogenous relationship between sales and ratings. In other words, a "good" product is likely to experience higher sales and receive more positive ratings than a "bad" product. The consequence is that sales and ratings are correlated, but the relationship is not necessarily causal. Chevalier and Mayzlin (2006) control for product heterogeneity by comparing differences in sales ranks for a sample of books that sold on both Amazon.com and Barnesandnoble.com. Because each site operated independently, each had a different set of posted ratings. Their results indicate that, across the products in their sample, the valence and volume of posted ratings had significant effects on sales performance.

Overall, the existing research has identified a number of important ratings metrics that can influence sales. However, although there is ample research on the effects of ratings on sales, the understanding of the ratings behavior itself is relatively limited. Next, we discuss a few studies that have investigated ratings behavior and the social dynamics that have been observed.

Consumer Ratings Behavior

A few recent studies have found that posted online ratings exhibit systemic patterns over time; specifically, the valence of ratings tends to trend downward (Godes and Silva 2009; Li and Hitt 2008). Li and Hitt (2008) posit that this trend is part of the product life-cycle process, and as the product evolves, so does the customer base. Specifically, they argue that customers who buy early in the product life cycle have significantly different tastes and preferences from those who buy later. At the same time, initial product ratings tend to be provided by the early customers but consumed by the later customers; this results in an increasing level of dissatisfaction over time because potential buyers are reading the ratings and reviews of existing customers who have dissimilar preferences. Godes and Silva (2009) suggest an alternative explanation. They show that the valence of ratings decreases with the ordinality of the rating rather than time. The downward trend is explained by the decreasing ability of future buyers to assess similarity with past reviewers as the total number of ratings grows, which then leads to more purchase errors and, consequently, lower future ratings.

In an experimental setting, Schlosser (2005) demonstrates the effect of social influences on consumer rating behavior. She finds that consumers who have decided to post their opinions tend to negatively adjust their product evaluations after reading negative reviews, which indicates that consumer posting behavior is affected by social context and the valence of previously posted reviews. She attributes this behavior to the notion that posters strive to differentiate their reviews, and negative reviews are more differentiated because negative evaluators are perceived as more intelligent (Amabile 1983). This same mechanism may also be driving the downward trend in ratings that both Li and Hitt (2008) and Godes and Silva (2009) document.

Schlosser (2005) also discusses multiple-audience effects in the context of online posting behavior. Multiple-audience effects occur when people facing a heterogeneous audience adjust the message to offer a more balanced opinion (Fleming et al. 1990). This is yet another form of social influence and suggests that the effects of previously posted ratings on ratings behavior extend beyond the effect of valence and include the effect of variance.

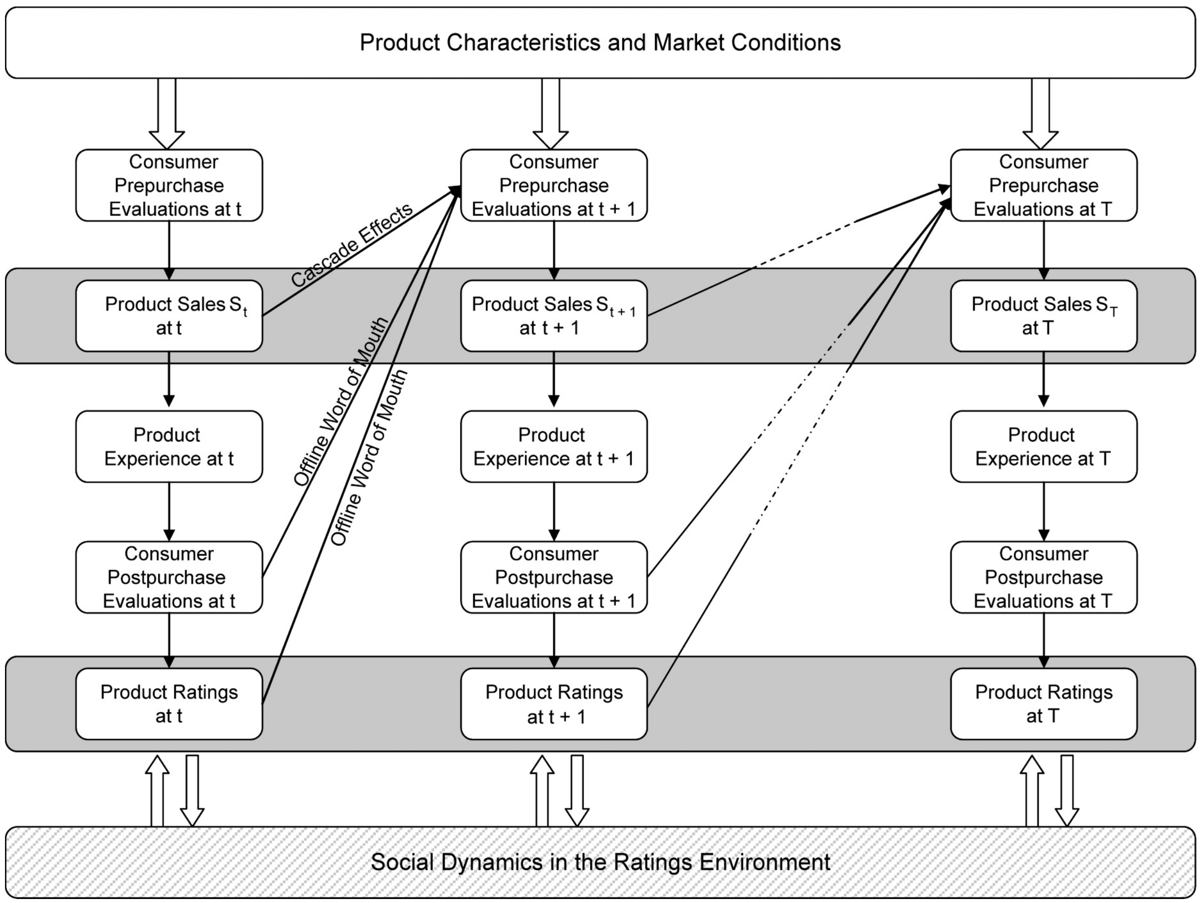
In this article, we consider potential social influences on ratings behavior by modeling the effects of both the valence and variance of previously posted reviews. We also allow for ratings dynamics by modeling the effects of the volume of posted ratings on subsequent rating behavior. By explicitly modeling these covariate effects, we can more effectively separate the effects of social influences and dynamics from the baseline ratings behavior.

CONCEPTUAL FRAMEWORK

Before presenting our data and proposed model, we discuss our conceptual framework. Figure 1 illustrates the relationship between product sales and ratings over time and incorporates key constructs from the consumer purchasing process. In this research, we observe product sales and product ratings in the data, and we model social dynamics in the ratings environment as a latent construct. We include the remaining constructs in the figure to complete the discussion of the consumer process.

Before any purchasing experience, a consumer constructs a prepurchase product evaluation based on a number of inputs. First, the consumer can independently evaluate

Figure 1
CONCEPTUAL FRAMEWORK



Notes: The constructs highlighted in the solid gray boxes are observed in our data, whereas the social dynamics construct highlighted in the textured gray box is modeled as latent. The empirical model proposed in this article focuses on the relationships between these three constructs.

product characteristics and market conditions to arrive at a purchasing decision. Second, the consumer can also look for signals of product quality in his or her social environment. Information cascade researchers suggest that consumers can be affected by the observable choices of others (Bikhchandani, Hirshleifer, and Welch 1992). These potential buyers can also be affected by the postpurchase product evaluations of other consumers through *offline* word of mouth (e.g., Westbrook 1987). In this article, we focus on the effect of online product ratings on these potential buyers through *online* word of mouth.

Consumer prepurchase product evaluations are related directly to product sales. After purchasing and experiencing the product, consumers update their postpurchase product evaluations with their own personal product experience. In theory, this postpurchase evaluation more heavily weighs the consumer's actual experience with the product and, as such, is more influenced by the consumer's independent assessment of the product and less influenced by the social factors that influenced the prepurchase evaluations.

In theory, these postpurchase evaluations should determine a consumer's posted product rating (if he or she chooses to post). However, other factors can also influence product ratings—namely, the social dynamics in the ratings environment. We posit that posted product ratings reflect the combined effect of (1) a consumer's socially unbiased postpurchase product evaluations and (2) any social dynamics that may influence a rater's public evaluation of the product in the ratings environment. In this article, we conceptualize the effect of social dynamics as the impact of previously posted ratings on the posting of future ratings. We do not differentiate between the influence of social dynamics on a potential rater's decision of whether to post and the decision of what to post. Instead, we examine the net effect of these dynamics on future posted ratings.

The objective of our modeling effort is to quantify these effects in terms of how they affect subsequent product sales. To that end, we develop a modeling framework that both captures the ratings process and measures the effect of ratings on product sales. In addition, we propose a set of metrics derived from our ratings model that enables us to decompose observed product ratings into components that represent (1) consumers' baseline (or socially unbiased) ratings for the product, (2) the effect of social dynamics in the ratings environment, and (3) idiosyncratic error. We present our model after discussing our data set in the next section.

DATA DESCRIPTION

We obtained our data from a national retailer of bath, fragrance, and beauty products and include a sample of 500 products rated and sold on the retailer's website. Products include hedonic items such as fragrances, room fresheners, scented candles, and bath salts as well as more utilitarian items such as skin care products (e.g., antiaging cream, moisturizer), sunscreen, and manicure/pedicure products. These items are moderately priced and appeal to the mass market (the highest-priced product in our sample is $25); they are not branded or designer products. The retailer in this study creates and produces its own products and sells them exclusively at its stores, both online and offline. Products tend to come in a high variety of fragrances. As a result, consumers frequently try new fragrances and/or include multiple fragrances of the same product in their purchase. The retailer does not engage in product-specific marketing. Instead, its marketing efforts are focused on promoting the entire store rather than individual products in the assortment. The wide variety of products, the inclusion of purely hedonic products (which are difficult to evaluate only on the basis of product attributes) and utilitarian products, and the absence of product-specific marketing activity make this an ideal data set for studying online product ratings and thus allow for the generalizability of our results across a number of different product types and contexts.

Our data span a one-year period from December 2006 to December 2007. Sales data are recorded weekly for each product. Although our data period includes two year-end shopping periods, the products we chose for our sample were nonseasonal items and did not display significant holiday sales patterns. The online-ratings functionality was introduced to the site in May 2007, halfway through the data period. The absence of a ratings tool on the site in the beginning of the study enables us to more accurately estimate a sales baseline for each product and thus separate the effects attributed to the rating system. During the postratings period, there was a sale event that affected some of our sample for two weeks. We control for this event in our analysis.

The ratings tool enabled customers to post both a star rating (on a five-star scale) and review text.[2] Each posting (rating and review) is recorded individually in our data set with a date stamp. Visitors to the website are presented with the average and number of ratings posted for the product being viewed. The entire history of previously posted ratings is also available.

Of the 500 products in our sample, 3801 ratings were posted. At the time we collected the data, the retailer made no promotional efforts to solicit the posting of online ratings. The ratings posted on this site are typical of most ratings posted online in that they are predominantly positive, a pattern consistently identified in previous research (Chevalier and Mayzlin 2006; Dellarocas 2003; Resnick and Zeckhauser 2002). In our sample, 80% of all product ratings were five-star ratings. If we examine the valence of ratings at the product level, 91.8% of the products in our sample received at least one five-star rating. Products also received their fair share of negative and/or neutral ratings. Table 2 provides a histogram of the proportion of products that received one-, two-, three-, four-, and five-star ratings.

This distribution of ratings results in a fair degree of ratings variance within products, with the average product experiencing a variance of .44 across ratings. Table 3 provides a brief description of the heterogeneity across products in our data set.

MODELING OVERVIEW

Our modeling objective is to separate the effects of social dynamics from those resulting from the consumers' socially unbiased product evaluation. We refer to the latter effect as the baseline ratings behavior and model how social dynamics can cause ratings to deviate from this baseline. We then model the effects of these socially influenced deviations on product sales.

First we develop a ratings model that is intended to decompose observed ratings into a baseline component and a social influence effect. A challenge we face in modeling the ratings process is that we must capture both what was posted and when it was posted, which enables us to differentiate between a product that received a five-star average rating but was sparsely rated to a more heavily rated product with the same average rating. To address this issue, we adopt a hazard modeling approach that treats the arrival of ratings of each star level as separate (though potentially correlated) timing processes. A well-liked product would tend to receive five-star ratings at a more frequent rate than one-star ratings; this would be discernable from the proposed hazard modeling framework.

We include time-varying covariates in each of the hazard processes to capture the effects of social influence and to control for potential dependencies across different star-level arrivals. The resulting baseline hazard rates would then describe the baseline ratings behavior and can be used to obtain a set of ratings metrics that represent the ratings that would have been received if not for the observed social dynamics in the ratings environment. Furthermore, this model enables us to obtain a set of expected ratings metrics that account for social dynamics by considering covariate effects.

Our ultimate goal here is to measure the sales impact of these dynamics. Therefore, we also present a sales model that captures the effect of ratings on product sales. Specifically, we decompose traditional ratings metrics into baseline ratings metrics, incremental effects resulting from social dynamics, and idiosyncratic error.

Table 2
HISTOGRAM OF RATINGS

Rating	*Number of Products*	*Percentage of Products*
Five stars	459	91.8
Four stars	213	42.6
Three stars	114	22.8
Two stars	82	16.4
One star	77	15.4

[2]In this study, we do not perform text analysis of the reviews. We recognize that text analysis may present another valuable dimension in explaining dynamics in ratings environments. We leave the exploration of review text for further research.

Table 3
DESCRIPTION OF DATA

	Mean	*Minimum*	*Maximum*
Ratings dates	—	May 31, 2007	November 29, 2007
Sales weeks (week ending)	—	January 6, 2007	December 8, 2007
Across Products			
Valence (average rating)	4.60	1	5
Variance	.44	0	4
Volume (number of ratings)	6.19	1	187
Total product sales	5449	51	36,999

A challenge in modeling the relationship between sales and ratings is the highly plausible endogenous relationship between the two. That is, a product may show higher sales not necessarily because of better ratings but rather due to a higher appeal to consumers, which is also reflected in higher ratings. This relationship between ratings and sales needs to be carefully addressed before any attributions can be made. We can address the potential endogeneity concern by leveraging the power of our unique data set, which includes weekly sales data both before and after the ratings functionality is available on the website. The observed sales data in the preratings period enable us to more effectively estimate a baseline level of sales for each product and to confidently attribute any postratings sales changes directly to the posted product ratings themselves.

RATINGS MODEL

Posted product ratings can be characterized by (1) the frequency of arrival and (2) the valence (or star level) of the ratings. To capture this process, we consider each star level separately and model the arrival of each of them as five parallel timing processes. Specifically, we assume that each rating level is associated with its own exponential hazard process (a process consistent with the observed data) with covariates.[3]

Because we observe the posting of multiple ratings, we write the hazard function governing the time until the next rating of the same valence (τ_{jvk}) as follows:

$$(1)\quad h(\tau_{vjk};\lambda_{vj},\boldsymbol{\beta}_{vj}) = \lambda_{vj}\exp\{\boldsymbol{\beta}_{vj}\mathbf{X}_{j\tau}\} \qquad \forall\ v \in \{1, 2, 3, 4, 5\},$$

where

j = product index,
k = index for rating occasion,
τ = time index in the ratings period (days), and
$\mathbf{X}_{j\tau}$ = vector of covariates.

We treat the posting of all ratings with valence (v) for a given product (j) as repeated events and model the time (τ) elapsing between the posting of rating k and rating k + 1 of the same valence. Each hazard function describes the frequency of posted ratings of that star level and decomposes it into a baseline hazard rate (λ_{vj}) and covariate effects ($\boldsymbol{\beta}_{vj}\mathbf{X}_{j\tau}$). The baseline hazard rate represents the underlying ratings behavior absent of any covariate effects. The β coefficients indicate how the specified covariates affect the rating frequency.

Note that this process is comparable to a Poisson model (which assumes exponential arrivals), in which we count the number of one-, two-, three-, four-, or five-star ratings in a given time period. However, the exponential hazard model is more precise because it examines the arrival time of each individual rating rather than aggregating ratings over a specified time period. Nonetheless, the interpretation of the parameters is the same.

We include a number of covariates that capture social influences and dynamics in the ratings environment. Because our context is one in which product-specific marketing is nonexistent, we do not include marketing-mix covariates. However, if necessary, these covariates would be easy to incorporate in our modeling framework for other contexts. One potential covariate of interest is price. Price can affect a consumer's assessment of the product's perceived quality and value (Dodds, Monroe, and Grewal 1991) and thus can affect ratings. Specific to our data set, price does not change over time[4]; thus, the effects of price can only be identified cross-sectionally (across products) but not temporally (within product). We accommodate for variations in price through the baseline effects.

For our analysis, we specify the following covariates:

$\text{LAGVALENCE}_{j\tau}$ = average of all ratings for j posted before τ (mean centered),

$\text{LAGVARIANCE}_{j\tau}$ = variance across all ratings for j posted before τ (mean centered),

LAGVOLUME_{j} = total number of ratings for j posted before τ (mean centered), and

SALE_{τ} = indicates a sale event at time τ.

The first three covariates (LAGVALENCE, LAGVARIANCE, and LAGVOLUME) characterize previously posted ratings and are metrics that are mirrored later in the sales

[3]Alternatively, the product-ratings arrival can be modeled in a count data framework by using the number of posts per unit of time as a dependent variable. We found the count data approach to be less appropriate in our case, because it is not obvious what level of (time) aggregation should be used in the multiproduct environment considered in this study. For some products with infrequent review posts, aggregation on a weekly (or even biweekly) level might be necessary to reliably estimate parameters of a count model, whereas for others (with a high level of post activity), weekly aggregation might be suboptimal and would effectively lead to information loss.

[4]An exception is the sale event, for which we control.

model.[5] These covariates are updated with the arrival of each new rating (of any valence).

The LAGVALENCE covariate enables us to observe how the positivity or negativity of past reviewers can influence future postings. On the basis of previous research, we expect that LAGVALENCE will have a negative effect on the subsequent arrival of negative ratings. The LAGVARIANCE covariate can capture potential multiple-audience effects. High LAGVARIANCE would indicate disagreement among reviewers and the presence of multiple audiences, possibly leading to more balanced and neutral future ratings (Fleming et al. 1990). We model dynamics with the LAGVOLUME covariate, which is similar to the way Godes and Silva (2009) model dynamics and the effect of ordinality. This covariate enables us to capture the trend in ratings as more ratings are posted. In addition, because of the parallel hazard functions, this covariate would enable us to identify the source of the downward trend. That is, is the downward trend in ratings driven by the posting of more negative ratings ($\beta_{\text{negative, LAGVOLUME}} > 0$), fewer positive ratings ($\beta_{\text{positive, LAGVOLUME}} < 0$), or both? Note that these covariates are time varying and continue to be updated as ratings of any valence gets posted. Therefore, to accommodate time-varying covariates (measured in discrete time), the pdf and survival function resulting from the hazard function specified in Equation 1 are as follows:

$$\text{(2)} \qquad f(\tau_{vjk}) = \lambda_{vj} e^{\beta_{vj} X_{j\tau}} \exp\left\{ -\lambda_{vj} \sum_{\substack{u = 1+ \\ \tau_{vj(k-1)}}}^{\tau_{vjk}} e^{\beta_{vj} X_{ju}} \right\}, \text{ and}$$

$$\text{(3)} \qquad S(\tau_{vjk}) = \exp\left\{ -\lambda_{vj} \sum_{\substack{u = 1+ \\ \tau_{vj(k-1)}}}^{\tau_{vjk}} e^{\beta_{vj} X_{ju}} \right\}.$$

To account for product heterogeneity (and the likely correlation between ratings arrivals of different star levels), we assume that baseline hazard rates, λ_{vj}, are drawn from a multivariate normal distribution as follows: $\ln \boldsymbol{\lambda_j} \sim \text{MVN}(\boldsymbol{\mu_\lambda}, \Sigma)$.[6] We also allow for heterogeneous covariate effects such that $\boldsymbol{\beta_{vj}} \sim \text{normal}(\boldsymbol{\mu}_{\beta_{vj}}, \sigma^2_{\beta_{vj}})$.

[5]Another covariate that may be predictive of ratings arrival is lagged sales. Although we believe that for mature product categories lagged sales are unlikely to be a strong predictor, we tested model specifications with and without a lagged sales covariate. The difference in fit between the two models is minimal (deviance information criterion of 54,770.5 vs. 54,787.8 with and without lagged sales, respectfully), and the inclusion of lagged sales does not meaningfully change parameter estimates. Therefore, to avoid potential endogeneity problems, we omit lagged sales from the ratings model. We thank the associate editor for this suggestion.

[6]Alternatively, we can estimate a set of nonparametric (fixed-effects) baseline parameters for each product j. However, the sparsity of the data necessitates that we rely on Bayesian shrinkage to a common prior. For example, the maximum-likelihood fixed-effect estimate for the one-star hazard would be perfectly 0 for the 423 products that never received such a rating. Our approach also requires the assumption that the random effects are orthogonal to the covariates. To test the sensitivity of our results to this assumption, we have attempted an alternative specification that replaces LAGVOLUME with an independent and exogenous measure (see the Web Appendix at http://www.marketingpower.com/jmrjune11). The analysis shows that the predicted metrics resulting from the ratings hazard models are not significantly affected. We thank the associate editor for pointing this out.

Incorporating repeated ratings observations for each product and the discrete time nature of our data, we write the resulting likelihood function for product j conditional on $\boldsymbol{\lambda_j}$ as follows:

$$\text{(4)} \qquad L_j = \prod_{v=1}^{5} \left[S_v\left(T - \sum_{k=1}^{K_{vj}} \tau_{vjk} \right) \prod_{k=1}^{K_{vj}} p_v(t_{vjk}) \right]^{\delta_{vj}} S_v(T)^{1 - \delta_{vj}},$$

where T is the length of our observed ratings data, K_{vj} is the total number of ratings posted for product j of valence v, and $\delta_{vj} = 1$ if $K_{vj} > 0$ ($\delta_{vj} = 0$ otherwise). Finally, $p_v(\tau_{vjk})$ is the probability of a ratings arrival of star-level v in the discrete time period τ_{vjk} and is defined as the difference between the survival rates at $\tau_{vj\,(k-1)}$ and τ_{vjk}:

$$\text{(5)} \qquad p_v(\tau_{vjk}) = S_v[\tau_{vj(k-1)}] - S_v(\tau_{vjk}).$$

DECOMPOSING RATINGS METRICS

Our objective in modeling the arrival of posted ratings is to decompose observed ratings metrics and separate social dynamic effects from the product-specific baseline rating. To this end, we develop a set of baseline metrics representing the ratings that would have arrived in the absence of social dynamics. We then predict the expected ratings with social dynamic effects. Thus, the difference between these expectations and the baseline represents the effect of social dynamics on ratings arrivals. Any remaining difference between observed ratings and expected ratings is attributed to idiosyncratic error.

In the model we presented, the baseline hazard rates represent the consumer population's underlying propensity to post a one-, two-, three-, four-, or five-star rating for a given product. Because we mean-centered the covariates used in our model, the baseline hazards are given by $\lambda_{vj}\exp\{\boldsymbol{\beta_{vj} X_{j1}}\}$, which represents the initial hazard rate at $\tau = 1$ before any ratings arrive (with non-mean-centered data, this simplifies to λ_{vj}). These baseline hazards enable us to predict, for each product j, the ratings that would have been posted in the absence of social influences and dynamics and to calculate the associated ratings metrics (i.e., average rating, ratings variance, and number of ratings).

In an exponential hazard model, the hazard rate represents the expected number of arrivals in a single time period. As such, we can calculate the proportion of one-, two-, three-, four-, or five-star ratings, absent covariate effects, as a ratio of these baseline hazard rates:

$$\text{(6)} \qquad q_{vj} = \frac{\lambda_{vj}\exp\{\boldsymbol{\beta_{vj} X_{j1}}\}}{\sum_v \lambda_{vj}\exp\{\boldsymbol{\beta_{vj} X_{j1}}\}}.$$

The expected average rating, $\hat{R}_j$, and the expected ratings variance, $\hat{V}_j$, for product j would then follow:

$$\text{(7)} \qquad \hat{R}_j = \sum_v (q_{vj} \times v), \text{ and}$$

$$\text{(8)} \qquad \hat{V}_j = \sum_v \left[q_{vj} \times (v - \hat{R}_j)^2 \right].$$

The number of ratings (regardless of valence) expected in the absence of social dynamics would simply be the total baseline hazard rate multiplied by time, τ. Note that unlike the metrics for average rating and ratings variance, we expect the baseline number of ratings to change from period to period as the length of the observation period increases:

$$\hat{N}_{j\tau} = \tau \times \sum_{v} \lambda_{vj} \exp\{\boldsymbol{\beta_{vj} X_{j1}}\}. \tag{9}$$

In addition to the baseline ratings metrics described in Equations 7–9, we can also calculate the metrics associated with a ratings process that is subject to social dynamics. Specifically, we use the time-varying hazard rates to predict the total number of one-, two-, three-, four-, and five-star ratings posted, $m_{vj\tau}$, as follows:

$$m_{vj\tau} = \sum_{u=1}^{\tau} \lambda_{vj} \exp\{\boldsymbol{\beta_{vj} X_{ju}}\}, \tag{10}$$

where the vector of covariates includes only LAGVALENCE, LAGVARIANCE, and LAGVOLUME. From this, we can easily compute the average rating ($R_{j\tau}$), ratings variance ($V_{j\tau}$), and number of ratings ($N_{j\tau}$) for product j at time τ as follows:

$$R_{j\tau} = \frac{\sum_{v} (m_{vj\tau} v)}{\sum_{v} m_{vj\tau}}, \tag{11}$$

$$V_{j\tau} = \frac{\sum_{v} \left[m_{vj\tau} (v - R_{j\tau})^2 \right]}{\sum_{v} m_{vj\tau}}, \text{ and} \tag{12}$$

$$N_{j\tau} = \sum_{v} m_{vj\tau}. \tag{13}$$

The difference between these metrics that account for social dynamics and the baseline metrics specified in Equations 7–9 would describe the impact of social dynamics on product ratings, whereas the difference between observed ratings metrics and the predictions from Equations 11–13 would be attributed to idiosyncratic error. We consider the impact of each ratings component on product sales in the "Sales Model" section.

SALES MODEL

Existing research has typically modeled the relationship between sales and ratings by regressing product sales against measures of ratings valence, variance, and volume. To be consistent with the existing research, we also focus on the valence, variance, and volume of posted product ratings. However, unlike existing research, we decompose these metrics into a baseline component, a social dynamics component, and an error component as described previously.

In addition, we divide our data into a preratings period and a postratings period. In the preratings period ($t < t^*$), we estimate only baseline sales for a given product, c_j, with no covariate effects. In the postratings period ($t > t^*$), we include a set of covariates (Z) that capture the effect of ratings on sales:

$$\ln(S_{jt}) = \begin{cases} c_j + \varepsilon_{jt} & \text{for } t < t^* \\ c_j + b_j Z_{jt} + \varepsilon_{jt} & \text{for } t > t^* \end{cases}, \tag{14}$$

where t indexes time (in weeks) for the entire data period and $\varepsilon_{jt} \sim \text{normal}(0, \sigma_\varepsilon^2)$.

The first covariate we include is a postratings indicator variable to control for the effect of introducing the ratings functionality to the site (POSTRATING_t). We also include an indicator variable to control for the effect of a sale event that occurred in the post-ratings period (SALE_t). To capture the effects of the posted ratings themselves, we separately consider the baseline, social dynamics, and error components of each of our three ratings metrics.

The baseline ratings metrics include the mean-centered average rating, $\hat{R}_j$ the mean-centered ratings variance, $\hat{V}_j$; and the number of ratings, $\hat{N}_{jt}$. Because the baseline average rating and ratings variance are product specific and do not vary over time, we include the mean-centered values in our sales model to better reflect the impact of a product's baseline rating relative to the average product. To measure the effects of social dynamics on sales, we include the difference between the predicted rating metrics after adjusting for social dynamics (computed at the beginning of week t) and the baseline metrics for average rating ($R_{jt} - \hat{R}_j$), ratings variance ($V_{jt} - \hat{V}_j$), and the number of ratings ($N_{jt} - \hat{N}_{jt}$). We also include the *observed* deviations from predicted average rating ($\text{AVGRATING}_{jt} - R_{jt}$), ratings variance ($\text{VARIANCE}_{jt} - V_{jt}$), and number of ratings ($\text{NUMRATINGS}_{jt} - N_{jt}$).

Finally, we accommodate unobserved product heterogeneity in the sales constant [$c_j \sim \text{normal}(m_c, s_c^2)$].[7] For completeness, we also model covariate effects to be heterogeneous across products according to the following: $b_{kj} \sim \text{normal}(m_{kj}, s_{kj}^2)$.

MODEL ESTIMATION AND RESULTS

We estimate our model using WinBUGS, specifying appropriate and diffuse priors. We ran at least 50,000 iterations, discarding the first 25,000 for burn-in. We used multiple starting values to test the sensitivity of the parameter estimates to starting values and to monitor convergence. The results indicate that starting values had no substantial impact on the parameter estimates. In addition, we computed Gelman–Rubin statistics for each parameter to monitor convergence.

We considered two approaches to model estimation: simultaneous and two-stage estimation. The benefit of the simultaneous approach is that the uncertainty in parameter estimates in the rating model (as reflected in posterior distributions) is naturally incorporated into the sales model. The key downside is a significant computational burden. Therefore, we performed simultaneous and two-stage model estimations for a random subsample of 50 products and then compared posterior distributions of the key parameters of the sales model. Because we found no substantial difference

[7]We also considered a fixed-effects model. We found parameter estimation results for both specifications to be similar.

between the two approaches, we adopted the two-stage estimation approach.

Table 4 provides fit statistics for both the ratings and the sales models. Using our model estimates, we compute the pseudo-R-square value (i.e., squared Pearson correlation between observed and predicted values) and the mean absolute percentage error for average rating, ratings variance, number of ratings, and product sales. Overall, both the ratings and the sales models fit our data quite well.

Ratings Model Results

Table 5 provides the parameter estimates resulting from our proposed ratings model.[8] The results show that ratings dynamics can substantially affect the arrival of future ratings through valence, variance, and volume effects. The LAGVALENCE coefficients indicate that increases in average ratings tend to encourage the subsequent posting of negative ratings ($\beta_{val,1}$ = .451, $\beta_{val,2}$ = .382, and $\beta_{val,3}$ = .568) and discourage the posting of extremely positive, or five-star, ratings ($\beta_{val,5}$ = –.0425). This result is consistent with the extant literature that provides evidence of differentiation behavior in the ratings environment (Amabile 1983; Schlosser 2005).

In addition, LAGVARIANCE has significant, negative effects on extremely negative ($\beta_{var,1}$ = –.302, $\beta_{var,2}$ = –.316) and extremely positive ($\beta_{var,5}$ = –.188) ratings; the effects on moderate three- and four-star ratings are insignificant. In other words, disagreement among raters tends to discourage the posting of extreme opinions by subsequent raters. This is consistent with the multiple-audience effects that indicate that consumers facing a highly varied audience are less likely to offer extreme opinions to avoid alienating any one segment of the audience (Fleming et al. 1990; Schlosser 2005).[9]

With respect to the LAGVOLUME effects, we observe that as the number of posted ratings increases, ratings of all star levels become more frequent. However, the magnitude of the volume effect on negative (one- and two-star) ratings is noticeably larger than that on more positive ratings, which suggests that although positive ratings may become more likely as ratings volume increases, they are overshadowed by the increased arrival of negative ratings. The net effect is a negative trend in posted product ratings. This result is consistent with the trends documented by Godes and Silva (2009), who show that the average rating decreases as the number of ratings increases. Our results add to their empirical findings and show that the decreasing trend in average ratings is driven by an increase in negative ratings (rather than a decrease in positive ratings, which would generate the same negative trend).

From the ratings-model results we have presented, we compute a set of metrics that decompose ratings into a baseline component, social dynamic effects, and idiosyncratic error for each product over time. Overall, the results from the ratings component of the model suggest that there are significant dynamics in ratings behavior. For each product, we compute the average difference between the socially unbiased versus socially adjusted ratings valence ($R_{jt} - \hat{R}_j$) variance ($V_{jt} - \hat{V}_j$), and volume ($N_{jt} - \hat{N}_{jt}$) and present the median, the 25th percentile, and the 75th percentile values (see Table 6). For the median product, observed social dynamics result in a lower average rating (–.0745), a greater variance among ratings (.133), and a lower volume of ratings (–.415). These results suggest that observed social dynamics seem to decrease ratings valence and volume while increasing ratings variance in our data.

Table 4
FIT STATISTICS

	Pseudo-R^2	*Mean Absolute Percentage Error*
Average rating (R)	.644	.0996
Ratings variance (V)	.757	.0771
Number of ratings (N)	.964	.189
Sales (S)	.998	.0583

Table 5
PARAMETER ESTIMATES FOR PROPOSED RATINGS MODEL

	M (μ)		*Variance^{-1} ($1/\sigma^2$)*
Baseline Rating Behavior			
$\ln\lambda_1$	–8.487	(.331)	
$\ln\lambda_2$	–8.208	(.400)	
$\ln\lambda_3$	–7.840	(.305)	
$\ln\lambda_4$	–5.899	(.183)	
$\ln\lambda_5$	–3.601	(.116)	
Effect of LAGVALENCE (β_{val})[a] on			
One-star ratings	.451	(.134)[b]	3.659 (1.508)
Two-star ratings	.382	(.134)[b]	3.588 (1.596)
Three-star ratings	.568	(.143)[b]	3.009 (1.328)
Four-star ratings	.0845	(.0582)	5.856 (1.867)
Five-star ratings	–.0425	(.0198)[b]	18.760 (2.624)
Effect of LAGVARIANCE (β_{var})[a] on			
One-star ratings	–.302	(.185)[c]	2.543 (1.196)
Two-star ratings	–.316	(.225)[c]	2.727 (1.351)
Three-star ratings	.00934	(.175)	2.927 (1.546)
Four-star ratings	.000713	(.108)	3.826 (1.516)
Five-star ratings	–.188	(.0671)[b]	5.565 (1.689)
Effect of LAGVOLUME (β_{vol})[a] on			
One-star ratings	.0853	(.0300)[b]	24.600 (4.1936)
Two-star ratings	.112	(.0296)[b]	25.62 (4.13)
Three-star ratings	.0531	(.0254)[b]	29.59 (4.275)
Four-star ratings	.0720	(.0187)[b]	44.28 (5.246)
Five-star ratings	.0478	(.0112)[b]	84.150 (7.226)
Effect of SALE$_t$ (β_{sale}) on			
One-star ratings	–1.598	(.736)[b]	1.02 (.822)
Two-star ratings	–.462	(.472)	1.566 (1.315)
Three-star ratings	–1.262	(.775)[b]	.672 (.973)
Four-star ratings	–.294	(.351)	1.032 (1.245)
Five-star ratings	.944	(.109)[b]	2.433 (.833)

[a]Measures are mean centered.
[b]Zero is not contained in the 95% confidence interval.
[c]Zero is not contained in the 90% confidence interval.
Notes: Values in parentheses represent standard errors of the estimates.

[8]We performed a set of robustness checks using both simulated data and subsampling techniques. The results of these tests confirmed a reasonably good performance of the proposed model. We report the details in the Web Appendix at http://www.marketingpower.com/jmrjune11.

[9]Although there is ample theoretical support for raters to exhibit differentiation behavior in response to ratings valence and variance on subsequently posted ratings, the nature of our secondary data does not allow us to rule out some alternative explanations. Unless rating manipulations are performed in a controlled experiment, the true causality is difficult to establish. We thank the associate editor for bringing this important point to our attention.

Table 6
EFFECTS OF SOCIAL DYNAMICS ON POSTED PRODUCT RATINGS

	25th Percentile	*Mdn*	*75th Percentile*
Valence $(R_{jt} - \hat{R}_j)$	–.143	–.0745	–.0287
Variance $(V_{jt} - \hat{V}_j)$	.0725	.133	.246
Volume $(N_{jt} - \hat{N}_{jt})$	–3.977	–.415	.0163

Although it is clear from our results that social dynamics can significantly influence subsequent rating behavior, the more managerially relevant question is: How do these dynamics affect sales? Therefore, we turn next to the results of the sales model and quantify the value of these dynamics in terms of product sales.

Sales Model Results

Table 7 presents the results of the proposed sales model, which decomposes ratings metrics into baseline, social dynamic, and error components. In Table 7, we also present the results for two benchmark models that measure the effects of observed average rating ($AVGRATING_{jt}$), ratings variance ($VARIANCE_{jt}$), and number of ratings ($NUMRATINGS_{jt}$) without any decomposition. In the first benchmark model (Benchmark 1), the product-specific sales constant is heterogeneous, but the covariate effects are not. In the second benchmark model (Benchmark 2), both the sales constants and the covariate effects are allowed to vary across products. The results show that Benchmark 2 (deviance information criterion [DIC] = 42,097.4) provides a significantly better fit than Benchmark 1 (DIC = 43,063.3), which suggests that there is heterogeneity in ratings effects across products.

When we decompose the ratings covariates in the proposed model, model fit further improves (DIC = 42,048.6), which indicates that the effects of product ratings on sales are more complex than what has been represented by the simple summary metrics in the benchmark models. The results of the proposed model, similar to those of the benchmark model, show significant effects of product ratings on sales. However, the baseline, social dynamic, and error components of the ratings metrics each have differing effects.

With respect to the valence of product ratings, we find that the baseline product rating has a positive impact on product sales ($b_1 = .577$). In addition, deviations from the expected baseline ratings caused by social dynamics $(R_{jt} - \hat{R}_j)$ have positive effects on product sales ($b_2 = 1.198$). This result suggests that positively (or negatively) valenced dynamics in the ratings environment can have direct effects on product sales.

The baseline variance also has a positive effect on sales ($b_4 = .464$), which suggests that products appealing to a broad base of customers with a high variety of opinions experience higher sales. However, our model results show that deviations from this baseline caused by social dynamics $(V_{jt} - \hat{V}_j)$ do not systematically affect product sales. Changes in variance caused by social dynamics may indicate that the dynamics are such that consensus opinions are encouraged (if $V_{jt} - \hat{V}_j < 0$) or that more varied opinions are encouraged (if $V_{jt} - \hat{V}_j > 0$). Although deviations in ratings variance do not have a significant direct effect on product sales, variance-related dynamics may affect future ratings, which in turn may affect long-term sales. We explore these potential indirect effects in the next section.

Our model results also indicate significant volume effects. It is not surprising that the baseline volume effect $(\hat{N}_{jt})$ is significant and positive ($b_7 = .0604$). When interpreting these results, it is important to note that unlike the measures for baseline valence and variance, the baseline volume metric does change over time and reflects the total expected number of posted ratings in each week (in the absence of social dynamics). Therefore, this result indicates the increase in sales resulting from the accumulation of posted product ratings over time. Because we also include a POSTRATING effect, this effect is above and beyond the sales increase resulting from the introduction of the ratings tool itself.

Overall, social dynamics have a direct effect on product sales through their effects on ratings valence. Although we have shown that social dynamics can also affect the variance and volume of ratings posted (see Table 5), our sales model results show that these influences do not have a direct (or immediate) effect on sales. However, because these dynamics do influence subsequent rating behavior, they may have an indirect (or longer-term) effect on sales. We examine these indirect sales effects in the next section.

THE IMPACT OF SOCIAL DYNAMICS ON FUTURE RATINGS AND SALES

In agreement with previous studies (e.g., Chevalier and Mayzlin 2006; Clemons, Gao, and Hitt 2006; Dellarocas, Zhang, and Awad 2007; Godes and Silva 2009; Li and Hitt 2008), our model results provide evidence that consumer-generated product ratings have a direct effect on product sales and are subject to social dynamics. The proposed modeling approach paired with a unique multiproduct longitudinal data set enables us to take the next step in this stream of research by answering the question of how social dynamics influence sales and how these effects can be quantified. The latter is of high interest to business practitioners who are eager to understand the effect of receiving a negative rating on product sales, the value of stimulating positive posts to rebut negative ones, and the long-term implications of these fluctuations in the rating environment. The comprehensive decomposition approach we developed in this article sheds light on these questions.

We illustrate the overall effects (both direct and indirect) of ratings dynamics on sales by simulating a number of scenarios for which the initial ratings for a given product are varied. Because rating behavior is sensitive to previously posted ratings, each scenario would generate a different dynamic in the ratings environment. Comparisons across these scenarios will enable us to examine the impact of these different dynamics on future ratings and sales. For this purpose, we select one product from our data set that, from the firm's perspective, can be considered to be representative,[10] and we simulate ratings and sales by using its product-specific posterior estimates.

[10]During the time period in question, this product had an average rating of 4.56, received a fair number of reviews (36 posts), and exhibited a notable ratings variance of 1.58.

Table 7
PARAMETER ESTIMATES FOR SALES MODELS

	Proposed Model		*Benchmark 2*		*Benchmark 1*
	Mean (m)	*Variance^{-1} (1/s^2)*	*Mean (m)*	*Variance^{-1} (1/s^2)*	*Parameter Estimate*
Mean sales constant (c_j)	3.255	.439	3.253	.437	3.253
	(.0687)	(.0287)	(.0674)	(.0284)	(.0688)
POSTRATING$_t$ (b_0)	.0273	7.364	.297	7.712	.143
	(.0673)	(1.642)	(.0389)[a]	(1.640)	(.0176)[a]
Sales Effect of Valence Metrics					
Observed					
AVGRATING$_{jt}$	—	—	.317	.881	.179
			(.102)[a]	(.218)	(.0349)[a]
Decomposed					
$\hat{R}_j$ (b_1)[b]	.577	4.201	—	—	—
	(.206)[a]	(1.69)			
$R_{jt} - \hat{R}_j$ (b_2)	1.198	1.925	—	—	—
	(.359)[a]	(1.263)			
AVGRATING$_{jt} - R_{jt}$ (b_3)	.0420	33.42	—	—	—
	(.0167)[a]	(3.901)			
Sales Effect of Variance Metrics					
Observed					
VARIANCE$_{jt}$	—	—	.0422	5.221	.0842
			(.0664)	(1.631)	(.0270)[a]
Decomposed					
$\hat{V}_j$ (b_4)[b]	.464	4.087	—	—	—
	(.222)[a]	(1.707)			
$V_{jt} - \hat{V}_j$ (b_5)	.2424	3.425	—	—	—
	(.254)	(1.466)			
VARIANCE$_{jt} - V_{jt}$ (b_6)	–.0437	4.181	—	—	—
	(.0688)	(1.208)			
Sales Effect of Volume Metrics					
Observed					
NUMRATINGS$_{jt}$	—	—	.0502	67.280	.00665
			(.00846)[a]	(5.901)	(.00196)[a]
Decomposed					
$\hat{N}_{jt}$ (b_7)	.0604	57.88	—	—	—
	(.0160)[a]	(6.251)			
$N_{jt} - \hat{N}_{jt}$ (b_8)	.0307	34.93	—	—	—
	(.0237)	(4.697)			
NUMRATINGS$_{jt} - N_{jt}$ (b_9)	–.0312	13.78	—	—	—
	(.0277)	(2.461)			
SALE$_t$ (b_{10})	.443	4.506	.441	4.592	.366
	(.0567)[a]	(1.192)	(.0512)[a]	(1.245)	(.0405)[a]
DIC	42,048.6		42,097.4		43,063.3

[a]Zero is not contained in the 95% confidence interval.
[b]Measures are mean-centered across the sample of 500 products.
Notes: Observed AVGRATING, VARIANCE, and NUMRATINGS covariates are mean centered in the benchmark model. Values in parentheses represent standard errors of the estimates.

Specifically, we simulate three scenarios. In all three scenarios, we hold constant the ratings volume in the initial month of the observation period. For the selected product, nine ratings of varying valence arrived in the first month. In Scenario 1, we manipulated the valence and variance of these nine ratings such that all were five-star ratings (“consensus five-star”); in Scenario 2, we consider a consensus three-star rating (“consensus three-star”); and in Scenario 3, we consider a ratings environment that averaged three stars but exhibited significant variation around this average rating (“varied three-star”).

Table 8 presents the simulated ratings metrics for the 3 months after the initial period in which ratings were manipulated. Across scenarios, ratings dynamics lead to virtually the same average rating after just a few months regardless of the valence or variance of initial ratings. In other words, in the absence of new external shocks to the system, average ratings converge to the product’s underlying baseline value over time. However, the nature of this convergence is a function of the initial ratings posted. Figure 2 plots the dynamics in ratings valence at the weekly level. Our model results indicate that higher average ratings increase the likelihood that one-, two-, and three-star ratings are subsequently posted. The impact of this dynamic is highlighted when comparing Scenario 1 (which has an average initial rating of five stars) to Scenarios 2 and 3 (which both have an initial average rating of three stars).

Comparing Scenarios 2 and 3 highlights the dynamics resulting from ratings variance. Note that in the varied three-star case, average ratings “recover” more quickly than in the consensus three-star scenario.

Likewise, the resulting variance in ratings also reveals a converging trend. Ratings variance in Scenarios 1 and 2 quickly increases from 0 to approximately .65 and remains at that level. However, the initially high variance in Scenario 3 steadily decreases throughout the simulation period.

Although ratings valence and variance both seem to be gravitating toward a baseline value, ratings volume exhibits

Table 8
SIMULATED RATINGS METRICS

	Scenario 1 (Consensus Five-Star)	*Scenario 2 (Consensus Three-Star)*	*Scenario 3 (Varied Three-Star)*
Average Rating[a]			
One month	5	3	3
Two months	4.64	3.97	4.17
Three months	4.62	4.56	4.66
Ratings Variance[a]			
One month	0	0	2.25
Two months	.57	.61	2.10
Three months	.63	.67	1.70
Ratings Volume[a]			
One month	9	9	9
Two months	13.65	12.67	13.48
Three months	18.00	16.83	18.35

[a]Ratings metrics refer to the average rating, ratings variance, and number of ratings at the start of the month.

more interesting differences across scenarios. Note that ratings volume is not manipulated in the initial setup of the simulation; thus, any differences across scenarios in terms of ratings volume is strictly a result of ratings dynamics. Note that the higher ratings variance in Scenario 3 seems to generate a slightly higher volume of ratings in the long run than the lower-variance scenarios.

These ratings dynamics are noteworthy in and of themselves. However, of particular managerial interest is the impact on product sales. Figure 3 plots these dynamics at the weekly level. It is not surprising that the initial positive ratings in Scenario 1 result in higher products sales than those in Scenario 2 (which has a lower initial average rating but the same ratings variance). As ratings valence begins to regress back to the product's baseline levels, sales decrease but increase again later as ratings volume increases.

Scenario 3 (varied three-star) provides another interesting illustration. Compared with the consensus three-star scenario (which has the same average rating but a lower ratings variance), product sales reach the level of the consensus five-star scenario more quickly when there is more variance in ratings. In contrast, it takes more than three months for sales in the consensus three-star condition to "recover" and reach the sales levels seen in the consensus five-star scenario.

This result, considered together with the demonstrated effects on ratings volume, indicates that variance in posted opinions can generate a dynamic that encourages subsequent word-of-mouth activity, which in turn facilitates not only the product's recovery in average ratings but also in product sales. The managerial implication of this result is that marketers should not necessarily encourage strictly positive word of mouth but also encourage variance of opinions in the online discussion.

Although the results of these simulations may be specific to the business environment of the collaborating firm, they should provide further impetus for models of consumer-generated product reviews to go beyond investigating the direct effects of ratings to incorporate a variety of potential indirect effects that come in through the dynamics studied here.

DISCUSSION AND CONCLUSION

Our objective was to decompose the ratings effect on sales into a baseline component and a dynamic component resulting from social dynamics in the ratings environment. We model ratings behavior as a dynamic hazard process and measure the effects that previously posted ratings have on future ratings behavior. The hazard modeling framework provided measures of expected average product rating, ratings variance, and ratings volume both with and without the influence of social dynamics. These metrics, when included in a model of product sales, enable us to measure the impact of ratings dynamics on product sales. Our model results and simulations show that there are substantial ratings dynamics, and their effects on sales are noticeable. Specifically, our model results show that ratings dynamics can have a direct effect on product sales through their impact on ratings valence but not variance or volume. Our simulation results further indicate that differences in initial ratings can result in dynamics that have indirect effects on product sales. However, these effects are relatively short lived.

In recent years, the sales impact of ratings has been the focus of many research efforts. However, few researchers have examined how ratings dynamics may influence product

Figure 2
AVERAGE RATING CONVERGENCE

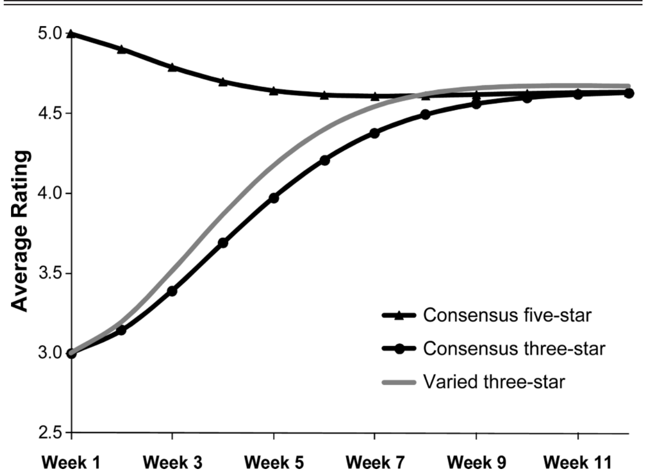
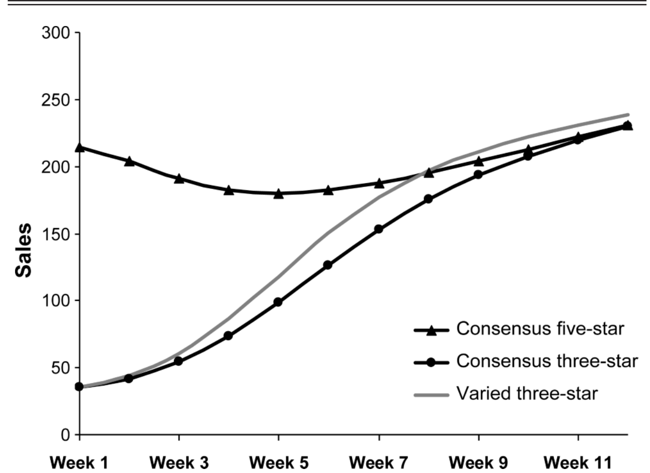


Figure 3
SALES DYNAMICS



sales. We have explicitly studied the ratings dynamics within a product forum and have measured the effects of product-level ratings measures on future ratings behavior. We hope that our results encourage further study of how people are influenced by the social dynamics taking place in a ratings environment. For example, it might be of interest to study whether social dynamics can affect a person's decision to participate in the ratings forum at all, potentially leading to systematic shifts in the customer base composition over time. Individual-level data would be helpful in addressing this question.

Furthermore, we constructed a model that allows for heterogeneous ratings effects across products. Although our results indicate that products do vary in terms of how they respond to ratings dynamics (in both the ratings and sales models), examining the sources of these variations is outside the scope of this article. We encourage researchers to study these differences across products. Furthermore, the analysis of heterogeneity across products can be improved if more ratings data on a product level become available. Indeed, because our data set is sparse (2500 transition processes are being estimated on the basis of 3801 transitions, and only 945 of 2500 processes are not right censored), we need to rely on Bayesian shrinkage estimation to infer parameters of each transition process. Richer data sets would allow for nonparametric baseline estimates, which may offer further insights into across-product variations.

Another important dimension of ratings environments that calls for further inquiry is the effects of variation in price and sales volume. In our data set, price does not change over time, precluding us from studying any temporal effects that may be caused by price fluctuation. Furthermore, we assume in our analysis that variations in ratings arrival rates are not a function of variations in previous sales. Although in our specific business setting we find little evidence that previous sales help predict future ratings, we acknowledge that this might be specific to mature product categories such as the ones studied here, and future studies should further investigate the role of previous sales on ratings. To establish the true causality of the effects studied here, further research may need to go beyond secondary data analyses and explore rating dynamics in controlled experiment settings.

Overall, online product ratings represent one type of online user-generated content, and as researchers, we have little understanding of the behavior driving consumers to provide this content or their responses to content provided by others. Although substantial research is still needed in this area, we hope that this article provides a first step in framing the problem, introducing a modeling approach, and presenting some empirical results that not only answer some questions but also stimulate readers to ask new ones.

REFERENCES

Amabile, Teresa M. (1983), "Brilliant but Cruel: Perceptions of Negative Evaluators," *Journal of Experimental Social Psychology*, 19 (2), 146–56.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch (1992), "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades," *Journal of Political Economy*, 100 (5), 992–1026.

Chen, Yubo and Jinhong Xie (2008), "Online Consumer Review: Word-of-Mouth as a New Element of Marketing Communication Mix," *Management Science*, 54 (3), 477–91.

Chevalier, Judith and Dinal Mayzlin (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 43 (August), 345–54.

Clemons, Eric K., Guodong Gao, and Lorin M. Hitt (2006), "When Online Reviews Meet Hyperdifferentiation: A Study of the Craft Beer Industry," *Journal of Management Information Systems*, 23 (2), 149–71.

Dellarocas, Chrysanthos (2003), "The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms," *Management Science*, 49 (10), 1407–24.

——— (2006), "Strategic Manipulation of Internet Opinion Forums: Implications for Consumers and Firms," *Management Science*, 52 (10), 1577–93.

——— and Ritu Narayan (2006), "A Statistical Measure of a Population's Propensity to Engage in Post-Purchase Online Word-of-Mouth," *Statistical Science*, 21 (2), 277–85.

———, Xiaoquan Zhang, and Neveen F. Awad (2007), "Exploring the Value of Online Product Reviews in Forecasting Sales: The Case of Motion Pictures," *Journal of Interactive Marketing*, 21 (4), 23–45.

Dodds, William B., Kent B. Monroe, and Dhruv Grewal (1991), "Effects of Price, Brand, and Store Information on Buyers' Product Evaluations," *Journal of Marketing Research*, 28 (August), 307–19.

Duan, Wenjing, Bin Gu, and Andrew B. Whinston (2008), "Do Online Reviews Matter? An Empirical Investigation of Panel Data," *Decision Support Systems*, 45 (4), 1007–16.

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 (4), 593–609.

Godes, David and Dina Mayzlin (2004), "Using Online Conversations to Study Word-of-Mouth Communications," *Marketing Science*, 23 (4), 545–60.

——— and Jose Silva (2009), "The Dynamics of Online Opinion," working paper, R.H. Smith School of Business, University of Maryland.

Li, Xinxin and Lorin M. Hitt (2008), "Self-Selection and Information Role of Online Product Reviews," *Information Systems Research*, 19 (4), 456–74.

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

Moe, Wendy W. and Peter S. Fader (2001), "Modeling Hedonic Portfolio Products: A Joint Segmentation Analysis of Music CD Sales," *Journal of Marketing Research*, 38 (August), 376–85.

Resnick, Paul and Richard Zeckhauser (2002). "Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System," in *The Economics of the Internet and E-Commerce: Advances in Applied Microeconomics*, Vol. 11, Michael R. Baye, ed. Amsterdam: Elsevier Science, 127–57.

Sawhney, Mohanbir S. and Jehoshua Eliashberg (1996), "A Parsimonious Model for Forecasting Gross Box-Office Revenues of Motion Pictures," *Marketing Science*, 15 (2), 113–31.

Schlosser, Ann (2005), "Posting Versus Lurking: Communicating in a Multiple Audience Context," *Journal of Consumer Research*, 32 (2), 260–65.

Westbrook, Robert A. (1987), "Product/Consumption-Based Affective Responses and Postpurchase Processes," *Journal of Marketing Research*, 24 (August), 258–70.