

## **Measuring the Value of Social Dynamics in Online Product Ratings Forums**

Wendy W. Moe  
Associate Professor of Marketing  
Robert H. Smith School of Business  
University of Maryland  
3469 Van Munching Hall  
College Park, MD 20742  
(301) 405-9187  
[wmoe@rhsmith.umd.edu](mailto:wmoe@rhsmith.umd.edu)

Michael Trusov  
Assistant Professor of Marketing  
Robert H. Smith School of Business  
University of Maryland  
3454 Van Munching Hall  
College Park, MD 20742  
(301) 405-5878  
[mtrusov@rhsmith.umd.edu](mailto:mtrusov@rhsmith.umd.edu)

September 2009

The authors would like to thank the Marketing Science Institute for the research grant provided to this project. The author would also like to thank BazaarVoice for the data analyzed in this paper as well as the invaluable conversations surrounding this topic.

## **Measuring the Value of Social Dynamics in Online Product Ratings Forums**

### **ABSTRACT**

Extant research has shown that consumer online product ratings can significantly influence product sales. However, these ratings have also been shown to be subject to a number of social influences. In other words, posted product ratings not only reflect the customers' experience with the product but also reflect the influence of others' ratings. The objective of this paper is to model posted product ratings in an effort to measure the impact of the social dynamics that may occur in a ratings environment on both subsequent rating behavior as well as product sales.

Our modeling efforts are two fold. First, we model the arrival of product ratings and separate the effect of social influences from the underlying or baseline ratings behavior. Second, we model product sales as a function of posted product ratings. However, rather than simply modeling the effects of observed ratings, we decompose ratings into a baseline rating and the contribution of social influence. From this model, we can measure the overall sales impact resulting from observed social dynamics.

We show that ratings behavior is significantly affected by previously posted ratings. We further show that the effect on sales resulting from this social dynamic is significant but relatively small compared to the effect that ratings have when they represent an unbiased and independent evaluation of the product. With the increased popularity of online discussion and ratings forums, many marketers have been investing in efforts to moderate these conversations or to contribute comments of their own, effectively biasing the sentiments expressed in the online forum. Our results show that while these efforts can affect sales, their effects are limited and short-lived.

## **Measuring the Value of Social Dynamics in Online Product Ratings Forums**

### **Introduction**

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 has the potential to significantly affect product sales (Chevalier and Mayzlin 2006; Dellarocas, Zhang and Awad 2008; Clemons, Gao and Hitt 2006). In theory, online review forums facilitate the exchange of information and help consumers make more informed decisions. 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 which 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) showed that posted product ratings are influenced by social dynamics. Specifically, the rating an individual posts for a product is affected by previously posted ratings. Additionally, Godes and Silva (2006) demonstrate ratings dynamics that result in a negative trend in posted product ratings as the volume of postings increase. A similar trend has also been shown by Li and Hitt (2008).

The consequence of the ratings dynamics described above 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 quite disconcerting for product marketers, and as a result, many marketers are investing in activities intended to create a more favorable social dynamic for their products with the intention of boosting sales (Dellarocas 2006).

Our objective in this paper is to measure the value of the social dynamics found in online customer rating environments. We do this by explicitly modeling 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"<sup>1</sup> product evaluations). Accordingly, the proposed approach allows us to decompose the sales impact of ratings into fixed (or baseline) and dynamic components.

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 parallel hazard processes. This allows us to capture the timing and the valence of posted ratings simultaneously. Additionally, we include time-varying hazard covariates to capture the effect of social influence on ratings behavior. The resulting baseline hazard rates, after controlling for the covariate effects, provide measures of the underlying ratings behavior and allow us to estimate a set of ratings metrics that represent customer posting behavior absent of social influence. Deviations in posted product ratings from these baseline metrics represent the impact of social dynamics.

To capture the sale impact of social dynamics, we model sales as a function of both the baseline ratings metrics and the observed deviations in ratings behavior from this baseline. One unique aspect of our research that differentiates our work from previous studies is that we model

---

<sup>1</sup> In the subsequent presentation we use the term "unbiased" in a very narrow sense specific to the discussed context.

variation in sales over a large set of products over time. Many of the published studies in this area of research use publically available data on ratings and product sales rank at a fixed point in time, allowing for only cross-sectional analyses (see for example, Chevalier and Mayzlin 2006). In contrast, we use a data set obtained from an online retailer that contains longitudinal ratings and sales data. This allows 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 allows us to explicitly model product heterogeneity using Bayesian methods. This is important in our efforts to separate the causal effects that ratings have on sales from the correlated effects resulting from the potentially endogenous relationship between product sales and ratings.

Several other aspects of our data are worth mentioning. First, the functionality to post ratings on the site was introduced approximately half way through of our data period. The weekly sales levels observed prior to the introduction of the ratings functionality allow us to estimate a product's baseline sales level absent of any ratings effects. Second is the product category. Previous researchers have typically focused on short life cycle products such as movies and books, likely due to the available of data in these categories. However, books and movies are unique in that they have product life cycles that are both short and follow predictable exponential patterns (Sawhney and Eliashberg 1996, Moe and Fader 2001). These products experience the greatest level of sales (and ratings activity) immediately after launch. Very quickly after that, sales (and ratings activity) taper off dramatically. The danger of using such product categories is that results can be very sensitive to when in the product life cycle the researcher collects the data. In this paper, 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 more

easily attributed to changes in the ratings environment and are less likely to be a result of changes due to the natural progression of the product life cycle.

Our empirical results show that there are substantial social dynamics in the ratings environment, and the impact of these dynamics on product sales is significant. However, the overall effect of ratings on sales is still mostly driven by measures of baseline ratings behavior. The sales impact of social dynamics is small by comparison. This suggests that the real value of online customer product ratings lies in the facilitation of word-of-mouth and information exchange, and artificially inflating (or deflating) ratings have a limited impact.

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 shown how social influences and dynamics can affect the posting of product ratings. Following the literature review, we present a conceptual framework that relates product sales to consumer online ratings. We then describe the data used in this paper. In particular, we will highlight some of the characteristics of the product category featured in our data that make it well-suited for the research questions being addressed. We then develop the model and propose a set of metrics based on the ratings component of the model that allows us to decompose the ratings effect and measure the sales impact of social dynamics. We finally present the results and highlight some of the implications in the remaining sections.

## **Literature Review**

There exists a growing body of research in the marketing literature as well as the information technology literature that examines the effects of online word-of-mouth. These papers have considered various forms of online word-of-mouth, including user-provided ratings

and reviews, newsgroup postings, etc. While some have focused on measuring the effects of online word-of-mouth on performance measures (e.g., sales, growth, TV 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 most frequently represented by an average rating measure (Chevalier and Mayzlin 2006; Clemons, Gao and Hitt 2006; Duan, Gu and Whinston 2008; Dellarocas, Zhang and Awad 2008). It has also been represented by some measure of positivity (or negativity) of ratings (Chevalier and Mayzlin 2006; Liu 2006; Godes and Mayzlin 2004). The variance seen 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), while volume is most commonly represented by the number of postings.

While a number of papers 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 papers (Liu 2006; Duan, Gu and Whinston 2008) have examined the impact of posted user-reviews on movie box office sales. While the effects of valence and volume were modeled in both cases, only the volume of word-of-mouth was significant. Dellarocas, Zhang and Awad (2008) also studied the effects of online word-of-mouth on movie box office sales. However, rather than modeling weekly sales, they examined the diffusion of sales. In their case, they found that both valence and volume of online word-of-mouth had significant effects. Another study by Clemons, Gao and Hitt (2006) focused their attention on

the craft beer category and found that the valence and variance (but not volume) of ratings affected sales growth. In particular, they find that the valence of the top quartile of ratings has the greatest effect on predicting sales growth.

**Table 1. Literature Review**

Article	Product Category	Dependent Variable	Significant WOM effects
Liu (2006)	movies	sales	number of posts
Duan, Gu and Whinston (2008)	movies	sales	number of posts
Dellarocas, Zhang and Awad (2008)	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	TV viewership ratings	entropy of posts number of posts
Chevalier and Mayzlin (2006)	books	sales rank	average rating number of ratings
This paper	bath, fragrance and beauty products	cross product temporal variation in ratings and sales	static and dynamic effects of ratings

One important challenge that must be addressed when studying the sales effect of ratings is that of accommodating unobserved product heterogeneity and 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 when compared to a “bad” product. The consequence is that sales and ratings are correlated, but the relationship is not necessarily causal. Chevalier and Mayzlin (2006) controlled for product heterogeneity by comparing differences in sales ranks for a sample of books that sold on both Amazon and BarnesAndNoble.com. Since 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.

While the aforementioned studies focus on the impact of ratings on sales, Godes and Mayzlin (2004) also consider the impact of sales on ratings. Specifically, they model television viewership as a function of the valence, variance and volume of online newsgroup postings. At the same time, they model the number of online postings as a function of lagged viewership on the premise that one cannot comment on a television show if they did not previously see it. Their results indicate that while online word-of-mouth affects future viewership, it is also affected by past viewership.

Overall, the existing research has identified a number of important ratings metrics that can influence sales. The importance of accommodating unobserved product heterogeneity and the potentially endogenous relationship between sales and ratings is also highlighted. However, while there is ample research on the effects of ratings on sales, our 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 shown that posted online ratings exhibit systemic patterns over time, specifically the valence of ratings tends to trend downward (Li and Hitt 2008, Godes and Silva 2008). Li and Hitt (2008) posit that this is part of the product life cycle process, and as the product evolves, so does the customer base. Godes and Silva (2008) suggest an alternative explanation. They show that the valence of ratings decreases with the *ordinality* of the rating, which indicates the potential presence of social influences on rating behavior.

One study that has very clearly shown social influences in the context of online consumer ratings is that conducted by Schlosser (2005) in an experimental setting. Schlosser (2005) found that consumers who have decided to post their opinions tend to negatively adjust their product evaluations after reading negative reviews. This indicates that consumer posting behavior is affected by social context and the valence of previously posted reviews. She attributes this behavior to the fact that posters strive to differentiate their reviews, and negative reviews are more differentiated since negative evaluators are perceived as more intelligent (Amabile 1983). This same mechanism may also be driving the downward trend in ratings documented by both Li and Hitt (2008) and Godes and Silva (2008).

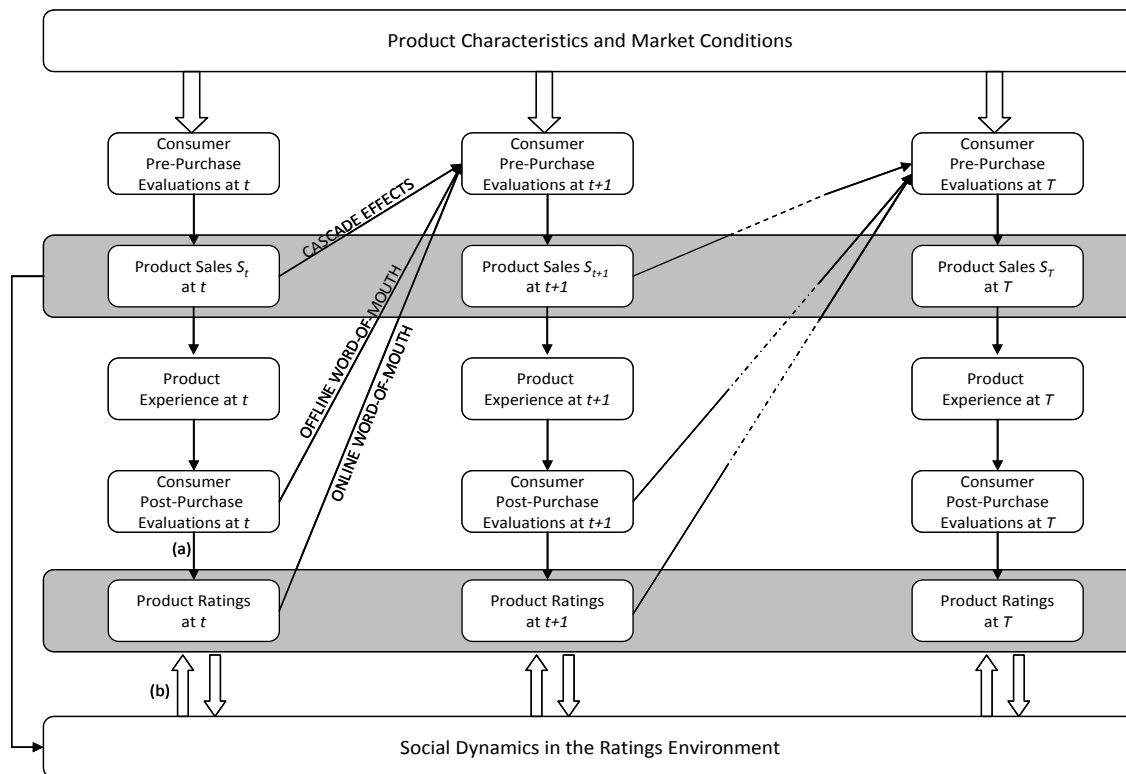
Schlosser (2005) also discusses multiple audience effects in the context of online posting behavior. Multiple audience effects occur when individuals facing a heterogeneous audience adjust the message to offer a more balanced opinion (Fleming et al. 1990). This is yet another form of social influence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

In this paper, 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, in this section 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.

**Figure 1.** *Conceptual Framework*



Prior to any purchasing experience, a consumer constructs a pre-purchase product evaluation based on a number of inputs. First, the consumer can independently evaluate product characteristics and market conditions to arrive at a purchasing decision. Second, the consumer can also look for signals of product quality in their 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

post-purchase product evaluations of other consumers via *offline* word-of-mouth (e.g., Westbrook 1987). Our focus in this paper is to examine the effect that online product ratings have on these potential buyers via *online* word-of-mouth.

Consumer pre-purchase product evaluations relate directly to product sales. Consumers who have purchased and then experienced the product then update their post-purchase product evaluations with their own personal product experience. In theory, this post-purchase 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 pre-purchase evaluations.

However, not all consumers with product experience choose to post a product rating. For those who do post, post-purchase evaluations can certainly influence the valence of their posted rating. But other factors can also influence product ratings, namely the dynamics in the ratings environment. Posted product ratings (and product sales<sup>2</sup>) can contribute to a social dynamic that can in turn affect future product ratings. As a result, product ratings are affected by both (a) post-purchase product evaluations and (b) social dynamics in the ratings environment, but the magnitude of each effect is unknown.

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. Additionally, we propose a set of metrics derived from our ratings model that allows us to decompose observed product ratings into components that represent (a) consumers' evaluation of product performance

---

<sup>2</sup> Cumulative product sales can influence the dynamics in a ratings forum since it affects the size of the potential rater population. Godes and Mayzlin (2004) show this effect in the context of TV viewership and discussion groups.

and (b) the effect of social dynamics in the ratings environment. We present our proposed model after discussing our data set in the next section.

## **Data Description**

Our data were obtained from a national retailer of bath, fragrance and beauty products and include a sample of 500 products rated and sold at their website. Products include hedonic items such as fragrances, room fresheners, scented candles, bath salts, etc. as well as more utilitarian items such as skin care products (e.g., anti-aging cream, moisturizer, etc.), sunscreen and manicure/pedicure products. Additionally, these are moderately priced items that appeal to the mass market (the highest priced product in our sample is \$25). These are not branded or designer products. The retailer in this study creates and produces their own products and sells them exclusively at their 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. Additionally, the retailer does not engage in product-specific marketing. Instead, their 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 based only on product attributes) as well as utilitarian products, and the absence of product specific marketing activity makes this an ideal dataset for studying online product ratings and thus allow for generalizability 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. While our data period includes two year-end shopping periods, the products chosen for our sample were non-seasonal items and did not display

significant holiday sales patterns. The online ratings functionality was introduced to the site in May 2007, half way through our data period. The absence of a ratings tool at the site in the beginning of the data allows us to more easily estimate a sales baseline for each product thereby separating the effects attributed to the rating system. During the post-ratings period, there was a sale-event that affected some of our sample for two weeks. We will control for this sale event in our analysis.

The ratings tool allowed customers to post both a star-rating (on a five-star scale) as well as review text.<sup>3</sup> Each posting (rating and review) is recorded individually in our data 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, a total of 3801 ratings were posted. At the time of the data, no promotional efforts were made by the retailer 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 did also receive their fair share of negative and/or neutral ratings. Table 2 provides a histogram of the proportion of products receiving a one, two, three, four or five star ratings.

---

<sup>3</sup> 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 rating environment. For example, Chen and Xie (2008) suggest that from online reviews consumer can learn whether and how a product matches a specific individual's preference and usage condition, which subsequently may influence purchase decision. Therefore, if consumers are better informed at the moment of making purchasing decision, then it is reasonable to expect that the level of satisfaction (post-purchase assessment) among informed buyers is higher compared to uninformed case. We leave the exploration of reviews text for future research.

**Table 2. Histogram of Ratings**

Rating	Number of products	Percent of Products
5-star	459 products	91.8%
4-star	213 products	42.6%
3-star	114 products	22.8%
2-star	82 products	16.4%
1-star	77 products	15.4%

This distribution of ratings results in a fair degree of ratings variance within products, with the average product experiencing a variance of 0.44 across ratings. Table 3 provides a brief description of the heterogeneity across products in our data set.

**Table 3. Description of Data**

	Mean	Minimum	Maximum
Ratings Dates		5/31/07	11/29/07
Sales Weeks (week ending)		1/6/07	12/8/07
<u>Across products:</u>			
Valence (Average Rating)	4.60	1	5
Variance	0.44	0	4
Volume (Number of Ratings)	6.19	1	187
Total Product Sales	55,589	51	36,999

## Modeling Overview

Our modeling objective is to separate the effects of social dynamics from that 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.

In this section, we first develop a ratings model that is intended to decompose observed ratings into a baseline component and a social influence effect. One challenge we face in modeling the ratings process is that we must capture both what was posted and when it was

posted. This allows us to differentiate between a product that received a 5-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 timing processes. A well-liked product would tend to receive 5-star ratings at a more frequent rate than 1-star ratings. This would be discernable from the proposed hazard modeling framework.

We include time-varying covariates to each of the hazard processes to capture the effects of social influence. 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. The difference between the observed ratings metrics and these derived metrics would then capture the overall effect of social dynamics.

The ultimate goal in this paper is to measure the sales impact of these dynamics. Therefore, in this section, we also present a sales model that captures the effect of ratings on product sales. Specifically, we decompose traditional ratings metrics into baseline ratings and incremental ratings resulting from social dynamics.

One 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 due to better ratings but rather due to a higher appeal to consumers, which is also reflected in higher ratings. This correlation between ratings and sales needs to be carefully addressed before any attributions can be made. Our modeling efforts address potential endogeneity concerns in several ways. First, our data set is unique in that it provides sales data both before and after the ratings functionality is available on the web site. The observed sales

data in the pre-ratings period allow us to more effectively estimate a baseline level of sales. This allows us to confidently attribute any post-ratings sales changes directly to the posted product ratings themselves. Second, the time series nature of the data allows us model the week-to-week variation in sales as a function of changes in the ratings environment. This provides a stronger test of the ratings effect on sales than a cross-sectional examination of sales and ratings. Finally, we explicitly model product heterogeneity using Bayesian methods to account for the differences in baseline sales across products.

### ***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 as five parallel timing processes. Specifically, we assume that each rating level is associated with its own exponential hazard process with covariates. Since we observe the posting of multiple ratings, we write the hazard function governing the time until the next rating of the same valence ( $t_{jvk}$ ) as follows:

$$(1) \quad h_v(t_{jvk}) = \lambda_{jv} e^{\beta_v \mathbf{X}_{jt}} \quad \forall \quad v \in \{1,2,3,4,5\}$$

where       $j$     = product index  
               $k$     = index for rating occasion  
               $\mathbf{X}_{jt}$  = vector of covariates

Each hazard function describes the frequency of posted ratings of that star-level and decomposes it into a baseline hazard rate ( $\lambda_{jv}$ ) and a vector of covariate effects ( $\beta_v$ ). The baseline hazard rate represents the underlying ratings behavior absent of any covariate effects. The  $\beta$ -coefficients indicate how the specified covariates affect the rating frequency.

We include a number of covariates that capture social influences and dynamics in the ratings environment. Because our context is one where product-specific marketing is non-existent, we do not include marketing mix covariates. However, these covariates would be very easy to incorporate in our modeling framework for other contexts, if necessary. One potential covariate of interest is that of price. Price can affect a consumer's assessment of the product's perceived quality and value (Dodds et al 1991) and thus can affect ratings. Specific to our dataset, price does not change over time, hence the effects of price can only be identified cross-sectionally (between products) but not temporally (with-in product). Since in the proposed model we control for product heterogeneity, a cross-sectional effect of price is captured by product-specific baseline effects.<sup>4</sup>

For our analysis, we specify the following five covariates:

$LAGVALENCE_{jt}$	= the average of all ratings for product $j$ posted <u>prior</u> to time $t$ .
$LAGVARIANCE_{jt}$	= the variance across all ratings for product $j$ posted <u>prior</u> to time $t$ .
$LAGVOLUME_{jt}$	= the total number of ratings for product $j$ posted <u>prior</u> to time $t$ .
$LNCUMSALES_{jt}$	= the sum of all product $j$ sales occurring <u>prior</u> to time $t$ .
$SALE_t$	= indicates a sale event

The first three covariates (LAGVALENCE, LAGVARIANCE and LAGVOLUME) characterize previously posted ratings and are metrics that are later mirrored in the sales model. These covariates are updated with the arrival of each new rating (of any valence).

The effects of social influence can be captured by the LAGVALENCE and the LAGVARIANCE covariates. The LAGVALENCE covariate allows us to observe how the positivity or negativity of past reviewers can influence future postings. Based on previous research, the expectation is that LAGVALENCE will have a negative effect on the subsequent

---

<sup>4</sup> As time-varying price covariates can easily be included in our model specification we believe that this might be an interesting research question to explore if the appropriate dataset becomes available.

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).

Dynamics are modeled by the LAGVOLUME covariate. This is similar to the manner in which Godes and Silva (2008) modeled dynamics and the effect of ordinality. This covariate allows us to capture the trend in ratings as more ratings are posted. Additionally, because of the parallel hazard functions, this covariate would allow 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 ( $\beta_{negative,LAGVOLUME} > 0$ ), fewer positive ratings ( $\beta_{positive,LAGVOLUME} < 0$ ) or both? The final covariate, LNCUMSALES, is designed to capture the effect of customer base size. As sales increase for a product, the number of potential posters increases. This should have a positive effect on the arrival of new product ratings. It is important to note that these covariates are time-varying and that they continue to be updated as ratings of any valence gets posted and sales accumulate. To accommodate time-varying covariates, the *pdf* and survival function resulting from the hazard function specified in equation (1) are as follows:

$$(2) \quad f_v(t_{jvk}) = \lambda_{jv} e^{\beta_v X_{jt}} \exp \left\{ - \lambda_{jv} \sum_{u=1}^{t_{jvk}} e^{\beta_v X_{ju}} \right\}$$

$$(3) \quad S_v(t_{jvk}) = \exp \left\{ - \lambda_{jv} \sum_{u=1}^{t_{jvk}} e^{\beta_v X_{ju}} \right\}$$

To account for product heterogeneity in baseline rating behavior, we assume that the baseline hazard,  $\lambda_{jv}$ , varies across products according to a gamma distribution with parameters  $r_v$  and  $\alpha_v$ .

$$(4) \quad g(\lambda_{jv}) = \frac{\alpha_v^{r_v}}{\Gamma(r_v)} \lambda_{jv}^{r_v-1} \exp\{-\lambda_{jv} \alpha_v\}$$

With repeated ratings for each product, the resulting likelihood function can be written as follows:

$$(5) \quad L_j = \prod_{v=1}^5 \int S_v(T - \sum_{k=1}^{K_{jv}} t_{jvk}) \prod_{k=1}^{K_{jv}} f_v(t_{jvk}) \cdot g(\lambda_{jv}) d\lambda_v$$

$$(6) \quad L = \prod_{j=1}^M L_j$$

where  $T$  is the length of our observed data,  $M$  is the total number of products,  $K_{jv}$  is the total number of ratings posted for product  $j$  of valence  $v$ .

### ***Baseline Ratings Metrics***

The primary objective in modeling the arrival of posted ratings is to obtain a measure of baseline rating behavior, absent of covariate effects. This is provided by  $\lambda_{iv}$ . The baseline hazard rates,  $\lambda_{iv}$ , provide the underlying propensity to post a 1, 2, 3, 4 or 5 star rating for a given product and allow us to predict the ratings that would have been posted in the absence of social dynamics. Specifically, we use the baseline hazard estimates to predict the underlying average rating, variance across ratings and the number of ratings for each product.

The baseline  $\lambda$  in an exponential hazard model represents the baseline rate of arrival, or number of arrivals expected in a single time period. From this, we can calculate the underlying proportion of 1, 2, 3, 4 or 5 star ratings as:

$$(7) \quad p_{vj} = \frac{\lambda_{vj}}{\sum_v \lambda_{vj}}$$

The expected average rating for product  $j$ ,  $\hat{R}_j$ , after removing covariate effects would then be:

$$(8) \quad \hat{R}_j = p_{1j} \cdot 1 + p_{2j} \cdot 2 + p_{3j} \cdot 3 + p_{4j} \cdot 4 + p_{5j} \cdot 5$$

Conceptually, this measure should reflect the consumer's independent and unbiased by social effects evaluation of product  $j$  (we will explore this relationship later in the paper), and in a world with no social dynamics, the observed average rating for product  $j$  should equal  $\hat{R}_j$ .

Therefore, any deviations from  $\hat{R}_j$  represent the net effects of social dynamics.

The expected baseline variance,  $\hat{V}_j$ , can also be computed in a similar manner:

$$(9) \quad \hat{V}_j = p_{1j}(1 - \hat{R}_j)^2 + p_{2j}(2 - \hat{R}_j)^2 + p_{3j}(3 - \hat{R}_j)^2 + p_{4j}(4 - \hat{R}_j)^2 + p_{5j}(5 - \hat{R}_j)^2$$

One interpretation of variance in a ratings context is that it represents the underlying preference heterogeneity among the raters. From the perspective of the potential buyer, this translates to uncertainty and can have significant effects on product sales. Social dynamics could potentially influence the variance of posted product ratings. The extent of this influence can be measured by examining the deviation between  $\hat{V}_j$  and the observed variance.

Finally, we can also measure the expected baseline number of ratings,  $\hat{N}_{jt}$ . This measure differs from the measures of average rating and ratings variance in that it changes from period to period. We calculate the expected baseline number of ratings,  $\hat{N}_{jt}$ , as the number of ratings arrivals expected regardless of valence,  $\sum_v \lambda_{vj}$ , multiplied by time,  $t$ .

$$(10) \quad \hat{N}_{jt} = t \cdot \sum_v \lambda_{vj}$$

These metrics and deviations from these metrics will be incorporated into our sales model which we discuss next. We should also note that the uncertainty in the posterior distributions of  $\hat{R}_j$ ,  $\hat{V}_j$  and  $\hat{N}_{jt}$  may vary from product to product and, in general, depends on the number of ratings each product receives (i.e., sample size). We discuss this issue in the model estimation 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 (as described in the previous section) and a social dynamics component. We quantify the social dynamics component by calculating the deviation between the observed ratings metrics and the expected baseline value calculated in the previous section.

Additionally, our data is divided into a pre-ratings period and a post-ratings period. In the pre-ratings period ( $t < t^*$ ), we estimate only baseline sales for a given product,  $c_j$ , with no covariate effects. In the post-ratings period ( $t > t^*$ ), we include a set of covariates ( $Z$ ) that capture the effects of (1) the baseline rating behavior for a given product and (2) the deviations observed from that baseline rating behavior. Also included as a covariate is an indicator variable to control for the effect of a sale event at the site.

$$(11) \quad \ln(S_{jt}) = \begin{cases} c_j + \varepsilon_{jt} & \text{for } t < t^* \\ c_j + bZ_{jt} + \varepsilon_{jt} & \text{for } t > t^* \end{cases}$$

where  $\varepsilon_{it} \sim \text{Normal}(0, s)$  and

$$(12) \quad Z_{jt} = \begin{bmatrix} \hat{R}_j - avg\hat{R} \\ \hat{V}_j \\ \hat{N}_{jt} \\ R_{jt} - \hat{R}_j \\ V_{jt} - \hat{V}_j \\ N_{jt} - \hat{N}_{jt} \\ SALE_t \end{bmatrix}$$

The first three covariates capture the effect of baseline ratings on sales. The first reflects the underlying average rating for the product ( $\hat{R}_j$ ). Specifically, we take the deviation of this measure from the average  $\hat{R}_j$  across all products in the sample ( $avg\hat{R}$ ). We also include measures of baseline variance,  $\hat{V}_j$  and number of ratings,  $\hat{N}_{jt}$ , as defined in (9) and (10) in the previous section. These covariate effects capture the effect that a ratings environment uninfluenced by social dynamics would have on sales. One interpretation of these effects is that they represent the sales impact resulting from the facilitation of word-of-mouth and the sharing of product experiences enabled by the online ratings tool. Notice that the baseline valence and variance covariates are product specific and do not vary over time.

In contrast to the baseline ratings covariates, the next set of three covariates represents the effect of social dynamics on sales. Again, we focus on the valence, variance and volume of ratings and model the observed deviations from the baseline metrics. Specifically, we consider the observed average rating ( $R_{jt}$ ), observed ratings variance ( $V_{jt}$ ) and the observed number of ratings ( $N_{jt}$ ) and measure the difference between these observed metrics and the associated baseline metric. Notice that these metrics vary over time,  $t$ , since new ratings are posted from week to week.

Finally, we accommodate unobserved product heterogeneity in the sales constant ( $c_j \sim Normal(\mu, \sigma)$ ). This is an important element of the modeling effort since obtaining an accurate baseline sales level is critical in our effort to separate the causal effects of rating on sales from the simple correlated relationship between sales and ratings. With an accurate measure of baseline sales, any variation in week-to-week sales activities can be confidently attributed to changes in the ratings environment. The availability of sales data prior to the introduction of the ratings functionality on the website facilitates this effort.

### **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. Multiple starting values were used 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. Additionally, Gelman-Rubin statistics for each parameter were computed to monitor convergence.

We considered two approaches to model estimation: simultaneous and two-stage estimation. The benefit of the former approach is that the uncertainty in the parameters 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 (e.g., for the entire sample of 500 products it would have taken a few months of CPU time to complete a single run). 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.<sup>5</sup> Since we found no substantial difference between the two, we adopted a two-stage estimation approach.

### *Ratings Model Validation*

Our objectives in testing the ratings component of our model are to assess (1) the value of incorporating hazard covariates to capture social dynamics, (2) the robustness of the parameter estimates in a variety of data conditions and (3) the ability of the covariates to separate social dynamics from an underlying ratings behavior that can be considered unbiased.

We begin by estimating only the ratings component of our model both with and without the social influence covariates. A comparison of the Deviance Information Criterion (DIC) shows that the inclusion of covariates significantly improves the fit of the model from 56,701 to 56,237. This suggests that there are substantial dynamics taking place in ratings behavior.

To test the robustness of our model, we conduct two tests. First, we used simulated data to examine the performance of our proposed model. Specifically, we are interested in how well our procedure recovers the true baseline hazard rates. The simulated datasets are designed to be in the stochastic neighborhood of the data used in our field application. The parameter estimates are consistent with the simulation parameters. Additionally, the expected baseline hazard rates for each star-rating level ( $\lambda_v$ ) are accurately recovered.<sup>6</sup> This adds confidence to our ability to provide an accurate assessment of the baseline ratings behavior using the baseline hazard rates ( $\lambda_v$ ) from the model.

For our second test of robustness, we estimate the ratings model on subsamples of the data. Specifically, we compare the results from modeling only those products with two, five or

---

<sup>5</sup> Due to space consideration we do not discuss the estimation results in depth. The detailed results are available from the authors upon request.

<sup>6</sup> The detailed results are available from the authors upon request.

ten plus ratings to those from modeling all products in our data set. The purpose of this test is to assess the ability of the model to accurately reflect ratings behavior even in sparse data environments. Among our 500 products, 368 had two or more ratings, 211 had five or more ratings and 115 had ten or more ratings. To compare the model results, we examine the correlation between the  $\hat{R}_j$ ,  $\hat{V}_j$ , and  $\hat{N}_j$  measures resulting from the estimation based on the complete data and the estimation based on a subset of the data (Table 4). Overall, the correlations are high, suggesting that the model is effective at capturing the baseline ratings behavior in a number of data contexts.

**Table 4.** *Correlation with Baseline Ratings Measures from Complete Data Sample*

Sample of products with...	$\hat{R}$	$\hat{V}$	$\hat{N}$
2+ ratings	.9865	.9865	.9945
5+ ratings	.9526	.959	.9603
10+ ratings	.9053	.9099	.9297

The final test of our ratings model examines the ability of our baseline ratings metrics to represent an unbiased measure of product experience. To do this, we compare the estimated  $\hat{R}_j$  to quality ratings provided by Consumer Reports. However, since the products included in our data are not products that are rated by third party experts, we collect a separate sample of products that are both rated by Consumer Reports and available on Amazon.com. The presence on Amazon.com provides the data on customer product ratings. We use these data to estimate our ratings model and obtain measures of  $\hat{R}_j$ , which are then compared the Consumer Reports' expert ratings.

Our sampling procedure was as follows. From Consumer Reports, we selected a group of product categories which, first, had a relatively large number of products rated and, second, were carried by Amazon.com. Next, we matched each individual product rated by Consumer Reports with a corresponding product on Amazon.com. Using the Amazon Web Services for all the products for which a match was established, we downloaded a complete history of user reviews and star ratings. Finally, we compared the average rating ( $R_j$ ) calculated individually for each product from the Amazon data to our estimated measure of  $\hat{R}_j$ . We found that for some product categories (e.g., noise canceling headphones or GPS) the average rating ( $R_j$ ) was very similar to  $\hat{R}_j$ . Since our objective was to establish if the proposed measure  $\hat{R}_j$  is a better predictor for product performance when compared to the average rating (which is a subject to social influences), we further examine only those product categories for which the average rating deviated from our measure of  $\hat{R}_j$ . Specifically, we dropped from further analysis the product categories for which the correlation between  $\hat{R}_j$  and  $R_j$  was greater than 0.8. For the remaining products, we regressed the product quality ratings obtained from the ConsumerReports.com first on average ratings and then on  $\hat{R}_j$ . The results of the regressions are presented in Table 5.

As Table 5 shows,  $\hat{R}_j$  is a significant predictor for product quality (as determined by Consumer Reports), while average rating is not. In other words,  $\hat{R}_j$  is a better predictor of Consumer Reports' quality ratings than a simple average rating, suggesting that removing social effects from product ratings makes it a better approximation of product quality. It is worth noting that the connection between Consumer Reports' quality ratings and the estimated measure of  $\hat{R}_j$  is not that strong ( $R^2$  is only 12%). This result is in line with numerous past studies which

argue that professional product reviews (e.g., Consumer Reports) tend to focus on product attribute information (“product-based”), while consumer reviews are more likely to reflect a product match with a specific individual’s preference (“user-based”) (e.g., Garvin 1984, Chen and Xie 2008). Therefore, the two can be seen as related but still complementary approaches to product assessment.

**Table 5.** *Connecting Consumer Reports Quality Ratings with Average Ratings and  $\hat{Q}_j$ .*

Regressing Product Quality on...		
Average Rating	3.09	(2.32)
Intercept	55.73	(8.76)*
$R^2$	0.048	
<hr/>		
$\hat{R}_j$	6.77	(3.02)*
Intercept	44.13	(10.41)*
$R^2$	0.125	

\* indicates covariate effects that are significant at p=0.05

### *Ratings Model Results*

Table 6 provides the parameter estimates resulting from our proposed ratings model. The first two rows in Table 6 present the estimates for the gamma parameters,  $r$  and  $\alpha$ , which describe the underlying distribution of baseline hazard rates for a given star-rating across products. The next three rows present the effects of social dynamics on ratings behavior. The results suggest that the valence, variance and volume of previously posted ratings each affect the posting of future ratings. The negative effect of LAGVALENCE on 4-star ratings ( $\beta_1 = -.086$ ) suggests that as the average rating increases for a product, 4-star ratings become less frequent. One explanation of this effect is that if a product appears to be very positively received, the value

of posting a moderately positive rating (4-stars) decreases since it neither adds to nor differentiates from the already posted ratings.

**Table 6. Parameter Estimates for Proposed Ratings Model**

	1-star	2-star	3-star	4-star	5-star
<b>Baseline Rating Behavior:</b>					
r	.14 (.027)	.15 (.027)	.21 (.033)	.49 (.056)	.96 (.068)
$\alpha$	18.99 (4.35)	21.33 (4.62)	23.06 (4.89)	31.52 (5.53)	33.15 (5.26)
<b>Ratings Dynamics:</b>					
LAGVALENCE <sub>jt</sub> ( $\beta_1$ )	-.0022 (.076)	-.091 (.065)	.040 (.058)	-.086* (.035)	.012 (.13)
LAGVARIANCE <sub>jt</sub> ( $\beta_2$ )	-.23* (0.11)	-.13 (.13)	.053 (.11)	.16* (.066)	-.029 (.033)
LAGVOLUME <sub>jt</sub> ( $\beta_3$ )	.039* (0.16)	.030* (.12)	.0024 (.010)	.0070 (.0052)	.0061* (.0010)
LNCUMSALES <sub>jt</sub> ( $\beta_4$ )	-.22* (.053)	-.16* (.047)	-.20* (.042)	-.12* (.029)	-.00015 (.019)
SALE <sub>t</sub> ( $\beta_5$ )	.0035 (.44)	.34 (.35)	-.096 (.33)	.46* (.19)	1.19* (.059)

The variance of previously posted ratings (LAGVARIANCE) also has a significant effect on subsequent posting behavior. While increased variance has a negative effect on the arrival of 1-star ratings ( $\beta_2 = -.23$ ), it has a positive effect on the arrival of 4-star ratings ( $\beta_2 = .16$ ). In other words, if there is disagreement in the ratings forum, individuals are less likely to contribute an extreme, negative opinion and more likely to contribute a moderate positive opinion. This is consistent with the multiple-audience effects that show that consumers facing a highly varied audience are more likely to offer more moderate opinions to avoid alienating any one segment of the audience (Fleming et al 1990).

Finally, we also see significant effects of ratings volume. As the number of posted ratings increase, we see both an increase in negative ratings (1- and 2-star ratings) and an increase in extremely positive ratings (5-star ratings). However, the magnitude of the effect on negative ratings is substantially larger than that on 5-star ratings. This suggests that while 5-star ratings may become more likely as ratings volume increases, it is overshadowed by the increased

arrival of negative ratings. The net effect is a negative trend in posted product ratings. This is consistent with the trends documented by Godes and Silva (2008). In their paper, Godes and Silva (2008) show that average rating decreases as the number of ratings increase. 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). One possible explanation for this effect is that as the number of ratings increase, consumers increasingly post ratings that differentiate their opinions from those previously posted, and negative ratings are more differentiated than positive ratings (Schlosser 2005). Our results pertaining to the increase in 5-star ratings confirm this dynamic since 5-star ratings are more differentiated as they tend to be more extreme.

In contrast to the effect of volume, the effect of cumulative sales (LNCUMSALES) reduces the number of negative ratings while having no significant effect on the posting of 5-star ratings. In other words, as the product builds its customer base, fewer ratings are posted, perhaps because the product is well established and consumers feel that their opinions would not add much to the overall discussion.

Overall, the results from the ratings component of the model suggest that there are significant dynamics in ratings behavior. Table 7 provides a histogram of these effects and shows the prevalence of these dynamics in our data set. A large majority of the products received more favorable ratings as a result of ratings dynamics. This leads us to ask: *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.

**Table 7. Histogram of Social Dynamic Effects on Average Rating**

---

---

$R_{jT} - \hat{R}_j$	Number of Products
-2 to -1	3
-1 to -0.5	1
-0.5 to 0	9
0 to 0.5	78
0.5 to 1	269
1 to 1.5	139
1.5 to 2	1

---

---

### *Sales Model Validation*

The key contribution of the sales component of our model is its ability to decompose the effect of ratings into an effect of baseline ratings and a separate effect of social dynamics in ratings. Therefore, a logical benchmark would be a model that does not decompose ratings but instead simply measures the effects of average rating, variance in ratings and the number of ratings. Table 8 provides these model comparisons and reports measures of DIC. These results show that decomposing ratings notably improves model fit as indicated by significantly stronger DIC.

### *Sales Model Results*

Consistent with previous research, we find that the valence and volume of ratings are related to product sales. What differentiates our work from previous studies is that we decompose these effects into a component that reflects the baseline effect of ratings and a component that represents the effect of ratings dynamics.

**Table 8. Parameter Estimates for Sales Models**

	Proposed Model	Benchmark Model
Constant ( $c_j$ )	3.25 (0.069)	3.22 (0.070)
<u>Sales Effect of Baseline Rating Behavior:</u>		
$(\hat{R}_j - avg\hat{R})$ or Average Rating ( $b_1$ )	.16* (0.040)	0.046* (0.0044)
$\hat{V}_j$ or Variance of Ratings ( $b_2$ )	-0.0087 (0.20)	0.0075 (0.014)
$\hat{N}_j$ or Number of Ratings ( $b_3$ )	0.018* (0.0020)	0.0025** (0.0015)
<u>Sales Effect of Ratings Dynamics:</u>		
$R_{jt} - \hat{R}_j$ ( $b_4$ )	.071* (0.0089)	--
$V_{jt} - \hat{V}_j$ ( $b_5$ )	.0062 (0.017)	--
$N_{jt} - \hat{N}_j$ ( $b_6$ )	-.017* (0.0025)	--
SALE <sub>t</sub> ( $b_7$ )	.58* (0.044)	0.39* (0.039)
$\sigma$	.40	0.43
DIC	42,882	42,985

\* indicates covariate effects that are significantly different from zero with 95% confidence.

\*\* indicates covariate effects that are significantly different from zero with 90% confidence.

With respect to the valence of product ratings, we find that the baseline product rating has a significant and large impact of product sales ( $b_1=0.16$ ). Because we estimate a product specific sales constant, this effect is separate from any endogenous relationship between sales and product performance. Instead, the effect of  $\hat{R}_j$  can be interpreted as the effect that online products ratings, in general, have on sales via the facilitation of word-of-mouth. For a “good” product, increased word-of-mouth will stimulate sales whereas for a “bad” product, increased word-of-mouth will depress sales. What is more interesting is the effect of ratings deviations from  $\hat{R}_j$ . As our ratings model has shown, posted online ratings can deviate substantially from the underlying expected  $\hat{R}_j$  as a result of observed dynamics in the ratings environment. The

impact of this deviation is captured by the coefficient for the  $R_{jt} - \hat{R}_j$  covariate. While the effect of this deviation is significant ( $b_4 = 0.07$ ), it is substantially smaller than the effect of the baseline ratings valence. A one-point improvement resulting from improved product quality/experience would result in a 17.79% increase in product sales while a one-point improvement resulting from ratings dynamics would result in a 7.39% increase in product sales (the average deviation from baseline ratings is +0.78). The implication of this result is that addressing customer concerns about the underlying product quality/experience itself has a bigger impact on sales than managing the dynamics in the ratings environment, not surprisingly. However, this is not to minimize the benefits to product sales that may result from a more positive ratings environment. A 7% impact on product sales resulting from social dynamics that are largely out of a marketer's control can be quite substantial and important.

Our model results also indicate significant volume effects. To better understand the results pertaining to ratings volume (i.e., Number of Ratings), we compare the parameter estimates to those resulting from the benchmark model (Table 8) with product level heterogeneity but no decomposition of ratings effects.

The benchmark model results show that the effect of ratings volume is statistically insignificant<sup>7</sup>. However, when we decompose the volume measure into the baseline ratings volume and the volume resulting from dynamics, we find significant effects. As to be expected, the baseline ratings volume has a significant and positive effect on product sales ( $b_3 = .018$ ). In other words, if a product's underlying quality is one that generates a high volume of ratings, then the product should also benefit from increased sales once consumers are able to share their experiences with one another through the newly introduced ratings tool. However, our results

---

<sup>7</sup> When we do not accommodate product heterogeneity, all ratings covariates are significantly different from zero with 95% confidence.

also show that increased ratings volume that result from dynamics taking place in the ratings environment do not increase sales and in fact hurts sales ( $b_6 = -0.017$ ). One possible explanation for this negative effect is that an unusually high number of ratings may indicate that an unusual (and likely negative) social dynamic may be taking place in the ratings environment<sup>8</sup>. This in turn negatively affects product sales. The implication of these results are that higher ratings volume can be related to higher sales if the ratings are earned as a result of having an intrinsically better product. Increased ratings generated as a result of dynamics in the ratings environment may actually signal a negative dynamic that can reduce product sales.

Overall, there is a fair degree of variation in terms of how much products are helped by ratings, specifically ratings valence. Of the 500 products in the sample, approximately 20% of the products do not experience any significant sales increase as a result of ratings valence, and many of these products actually experience a significant sales decrease. To further investigate the differences between products that benefit from ratings and those that do not, we decompose the total valence effect for each product into the effect of baseline valence,  $b_1 \hat{R}_j$ , and the effect of valence related dynamics,  $b_2 (R_{jT} - \hat{R}_j)$  where  $R_{jT}$  represents the average rating for product  $j$  at the end of our data period,  $T$ . For ease of presentation, Figure 2 presents the results by product quartiles where the quartiles are defined according to the total valence effect. That is, products in the top quartile benefit the most from ratings while products in the bottom quartile benefit the least (if they benefit at all).

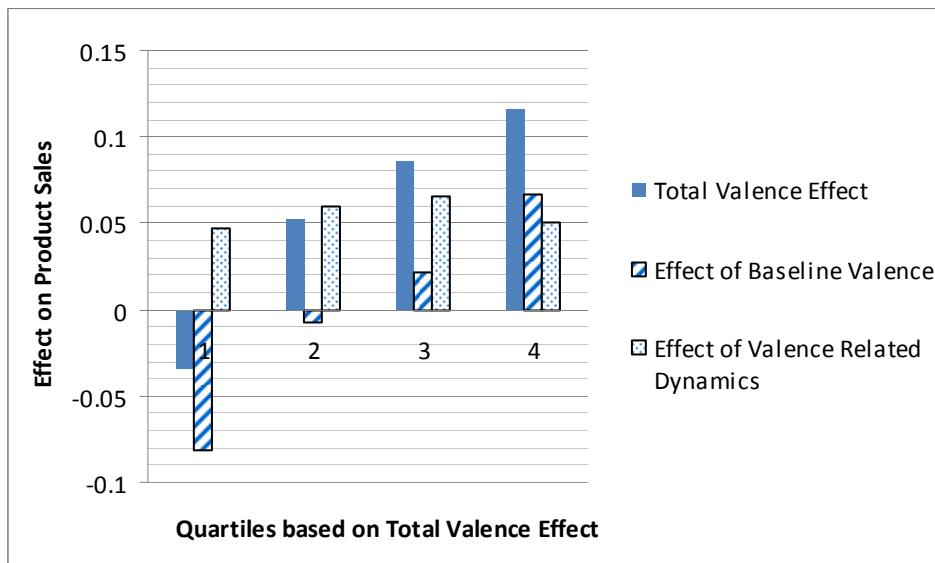
The results show that the differences in ratings effects across products are due primarily to baseline valence effects,  $b_1 \hat{R}_j$ , and not valence related dynamics,  $b_2 (R_{jT} - \hat{R}_j)$ . A comparison

---

<sup>8</sup> The correlation between the average ratings deviation,  $R_{jT} - \hat{R}_j$ , and number of ratings deviation,  $N_{jT} - \hat{N}_j$ , is -0.21

of effects across product quartiles illustrates this result. While the effect of baseline valence varies dramatically across product quartiles (and follows a trend similar to the total valence effect), the effect of valence related dynamics is relatively unchanged. This suggests that the primary difference between products that benefit from ratings and those that do not is underlying and unbiased consumer assessment of the product, as represented by our baseline valence metric. These results also suggest that for a fundamentally “good” product, online product ratings can facilitate positive word-of-mouth and increase product sales. However, for a “bad” product, online product ratings can alert consumers to the negative aspects of the product, thereby hurting sales.

**Figure 2.** *Decomposition of Total Valence Effect*



### **The Impact of Social Dynamics on Future Ratings and Long Term Sales**

Our model results provide evidence that ratings are subject to social dynamics, and these dynamics can affect product sales. However, like previous research, we have only demonstrated thus far the short term effects of ratings on sales. To investigate potential long term effects, we

simulate a variety of scenarios where we vary initial ratings for a given product and examine the resulting product sales. Since rating behavior is sensitive to previously posted ratings, each scenario would generate a different dynamic in the ratings environment.

We identified a number of “average” products from our sample based on their average rating and  $\hat{R}$  measure. We further identify a subset of products with a sufficient volume of ratings. From this subset, we selected one product that received both one- and five-star ratings early in the data period. Specifically, the product we used in our simulation had an average rating of 4.56 (the average for the entire product sample was 4.60) and  $\hat{R} = 3.75$  (the average for the entire product sample was 3.82). Additionally, this product received a total of 36 ratings in our data period resulting in a variance of 1.58. In the first month, this product received nine ratings which resulted in an average rating of 4.56 and a ratings variance of 1.78.

We simulate three scenarios and compare the resulting ratings and sales to the baseline scenario described above. In scenario one, we consider a situation where all nine ratings posted in the first month were 5-star ratings. In scenario two, we consider a consensus 3-star rating. Finally in scenario 3, we consider a ratings environment that again averaged 3-stars but exhibited significant variation around this average rating. Table 9 presents the simulation results.

Scenario 1 illustrates the outcome of a consensus 5-star rating. The increased average rating and the decreased variance results in a slight sales increase over the baseline sales levels. However, this increase decays over time as the average rating regresses to the baseline level and ratings variance increases.

Scenario 2 illustrates the effects of a consensus 3-star rating. Like in scenario 1, the average rating eventually approaches the average rating seen in the baseline scenario, despite the

poor initial ratings. Likewise, ratings variance also increases. As a result, the sales impact of a consensus 3-star rating is noticeably negative early on, but this effect fades over time.

Finally, scenario 3 demonstrates the effect of ratings variance. Like the other scenarios, the ratings metrics gradually return to those seen in the baseline scenario. As a result, any initial impact on sales eventually decays.

Overall, these simulations demonstrate that while ratings dynamics can have significant effects on product sales, these effects are short term. Over time, ratings and sales gravitate toward a baseline level that is indicative of the underlying consumer assessment of product performance.

**Table 9. Simulation Results**

	Baseline	Scenario 1	Scenario 2	Scenario 3
<u>Total Product Sales</u>				
1 month	740	758 (+2.5%)	658 (-11.2%)	660 (-10.9%)
2 months	1427	1443 (+1.1%)	1310 (-8.2%)	1322 (-7.4%)
3 months	2056	2064 (+.4%)	1933 (-6.0%)	1955 (-4.9%)
<u>Average Rating</u>				
1 month	4.56	5	3	3
2 months	4.55	4.58	4.37	4.36
3 months	4.49	4.48	4.48	4.50
<u>Ratings Variance</u>				
1 month	1.78	0	0	2.25
2 months	1.32	.58	.85	1.83
3 months	1.23	.80	.96	1.54

\* values in parentheses represent percent changes compared to baseline scenario.

## Discussion and Conclusion

The objective of this paper was to decompose the ratings effect on sales into a baseline component reflecting the consumers' unbiased evaluation of product quality 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 baseline measures of average product rating, ratings variance and ratings volume in the absence of social dynamics. These metrics, when included in a model of product sales, allow us to separate the effect of ratings as a signal of product quality from the effect of social dynamics in the ratings environment.

Our model results show that there are substantial ratings dynamics and that these dynamics do have a significant impact on product sales. However, the sales impact of ratings dynamics is substantially smaller than the impact associated with the baseline ratings metrics and the facilitation of word-of-mouth enabled by an online ratings tool. Our analysis shows that there is substantial variation in terms of how much ratings help product sales, and the magnitude of the ratings effect is largely driven by the underlying product evaluation. In our data, a one point increase in the baseline ratings metric is associated with a 17.79% increase in sales while a one point increase in ratings resulting from social dynamics is associated with a 7.39% sales increase.

Furthermore, our simulation demonstrates that any impact of ratings dynamics on product sales is short term and erodes over time. Our results show that while a few initial ratings can impact both product sales and ratings, these effects are short lived. The simulation results also show that over time, key ratings metrics begin to regress toward what is experienced in the baseline scenario. This leads to a decay in the sales impact.

Our results have important implications for our understanding of the how much online word-of-mouth can help a product. While positive online word-of-mouth may amplify the success of a fundamentally good product, it may do little to improve the performance of a bad

product. In the online environment, many marketers have invested in creating a positive ratings dynamic by moderating ratings forums, contributing their own comments, etc. Our findings suggest that this is not necessarily an effective strategy since the bulk of the ratings effect is tied to the underlying product evaluation and that any benefit achieved from these dynamics is likely short lived.

In recent years, the sale impact of ratings has been the focus of many research efforts. However, few have examined how ratings dynamics may influence product sales. In this paper, 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 the results of this paper encourage further study of how individuals are influenced by the social dynamics taking place in a product forum. For example, it might be of interest to study if a composition of buyer base changes over time as a result of rating dynamics. Related to this is the question of whether social dynamics can also affect an individual's decision to participate in the ratings forum at all. Individual level data would be helpful in addressing these questions.

Overall, online product ratings represent one type of online user-generated content, and as a field, we have very little understanding of the behavior driving consumers to provide this content or their responses to content provided by others. While substantial future research is still needed in this area, hopefully, this paper has provided a first step in framing the problem, providing 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 (March), 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).
- Chen, Yubo and Jinhong Xie (2008), "Online Consumer Review: Word-of-Mouth as a New Element of Marketing Communication Mix," *Management Science*, 54 (3), 477-491.
- Chevalier, Judith and Dinal Mayzlin (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 48 (August), 345-354.
- 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), 149-171.
- Congdon, Peter (2002), Bayesian Statistical Modelling, John Wiley & Sons, Ltd., Chichester, England.
- Dellarocas, Chrysanthos (2003), "The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms," *Management Science*, 49 (10), 1407-1424.
- Dellarocas, Chrysanthos (2006), "Strategic Manipulation of Internet Opinion Forums: Implications for Consumers and Firms," *Management Science*, 52 (10), 1577-1593.
- Dellarocas, Chrysanthos 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), 277-285.

- Dellarocas, Chrysanthos, Xiaoquan Zhang and Neveen F. Awad (2008), "Exploring the Value of Online Product Reviews in Forecasting Sales: The Case of Motion Pictures," *Journal of Interactive Marketing*, forthcoming.
- Dhar, Vasant and Elaine Chang (2007), "Does Chatter Matter? The Impact of User-Generated Content on Music Sales," *Working Paper*.
- 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*, Vol. 28, No. 3 (Aug., 1991), pp. 307-319.
- Duan, Wenjing, Bin Gu and Andrew B. Whinston (2008), "Do Online Reviews Matter? – An Empirical Investigation of Panel Data," *Journal of Retailing*, forthcoming.
- Erdem, Tulin, Michael P. Keane, and Baohong Sun (2008), "A Dynamic Model of Brand Choice When Price and Advertising Signal Product Quality," *Marketing Science*, 27 (6), 1111-1125.
- Fleming, John H., John M. Darley, James L. Hilton, and Brian A. Kojetin (1990), "Multiple Audience Problem: A Strategic Communication Perspective on Social Perception," *Journal of Personality and Social Psychology*, 58 (April), 593–609.
- Garvin, David A. (1984), "What Does 'Product Quality' Really Mean?" *MIT Sloan Management Review*, 26, no. 1, (Fall 1984).
- Godes, David and Dina Mayzlin (2004), "Using Online Conversations to Study Word-of-Mouth Communications," *Marketing Science*, 23 (4), 545-560.
- Godes, David and Jose Silva (2006), "The Dynamics of Online Opinion," *Working Paper*.

- Hu, Nan, Paul A. Pavlou and Jennifer Zhang (2006), "Can Online Reviews Reveal a Product's True Quality," *Proceedings of the Seventh ACM Conference on Electronic Commerce*, June, Ann Arbor, MI.
- Lai, DJ and Hardy, RJ (1999), "Potential Gains in Life Expectancy or Years of Potential Life Lost: Impact of Competing Risks of Death," *International Journal of Epidemiology*, 28 (5), 894-898.
- Li, Xinxin and Lorin M. Hitt (2008), "Self-Selection and Information Role of Online Product Reviews," *Information Systems Research*, 19 (4), forthcoming.
- 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 (3), 376-385.
- Narendranathan, Wiji and Mark B. Stewart (1991), "Simple Methods for Testing for the Proportionality of Cause-Specific Hazards in Competing Risks Models," *Oxford Bulletin of Economics and Statistics*, 53 (3), 331-340.
- Resnick, P. and Zeckhauser, R. (2002). "Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System." *The Economics of the Internet and E-Commerce*. Michael R. Baye, editor. Volume 11 of Advances in Applied Microeconomics. Amsterdam, Elsevier Science.
- Sawhney, Mohanbir S. and Jehoshua Eliashberg (1996), "A Parsimonious Model for Forecasting Gross Box-Office Revenues of Motion Pictures," *Marketing Science*, 15 (2), 113-131.

Schlosser, Ann (2005), "Posting Versus Lurking: Communicating in a Multiple Audience Context," *Journal of Consumer Research*, Vol. 32, September, pp. 260-265.

Westbrook, Robert A. (1987), "Product/Consumption-Based Affective Responses and Postpurchase Processes," *Journal of Marketing Research*, 24 (3), 258-270.