Pump up the Volume? Examining The Relationship Between Number of Online Reviews and Sales: Is More Necessarily Better?

Hila Etzion  
University of Michigan

Neveen Awad  
Wayne State University

Follow this and additional works at: http://aisel.aisnet.org/icis2007

Recommended Citation
http://aisel.aisnet.org/icis2007/120

This material is brought to you by the International Conference on Information Systems (ICIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in ICIS 2007 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
PUMP UP THE VOLUME? EXAMINING THE RELATIONSHIP BETWEEN NUMBER OF ONLINE REVIEWS AND SALES: IS MORE NECESSARILY BETTER?

Hila Etzion  
Ross School of Business,  
University of Michigan  
Ann Arbor, MI  
etzionh@umich.edu

Neveen Awad  
School of Business Administration  
Wayne State University  
Detroit, MI  
nawad@wayne.edu

Abstract

We study the relationship between volume of online product reviews and sales. We find that volume has a positive significant effect on sales of products whose valence (average rating) is perceived positively by consumers, while volume has a negative significant effect on sales of products whose valence is perceived negatively. Thus, pooling all products together in one regression model may not correctly capture the relationship between volume and sales, as the two opposing effects might cancel each other. In addition, we show that online reviews have a cross effect on sales of competing products: the demand for a product increases with its perceived quality, but decreases with the perceived quality of competing products. Finally, we show that consumers substitute the use of volume and valence when evaluating online review data: if the valence metric does not provide sufficient information to differentiate competing products, consumers would use the volume metric.

Keywords: Online Ratings, Online Word of Mouth, Product Reviews, Reviews Volume and Sales, Metrics of Online Reviews and Sales
### Introduction

Traditional attempts to measure word-of-mouth were based on two principal techniques: inference (Bass 1969; Reingen et al. 1984) and surveys (Bowman and Narayandas, 2001). In recent years, the growth of the Internet enables researchers to measure and watch the evolution of word-of-mouth directly and accurately from online review systems and feedback forums. Accordingly, researchers have been able to show that online product reviews (a specific form of online word-of-mouth) influence consumers’ purchase decisions and thus sales (Chevalier and Mayzlin, 2006; Senecal and Nantel, 2004). However, to date, the literature has been mixed in regards to which metrics of online product review data consumers use when making their purchase decisions.

Early research used *volume* and *entropy* when measuring the effect of online word-of-mouth on sales (Godes and Mayzlin, 2004). The theory behind volume is that the more consumers discuss a product, the higher the chance that other consumers would become aware of it and eventually buy it. The theory behind measuring entropy, or the spread of communication across communities, is that word-of-mouth spreads quickly within communities, but slowly across them (Granovetter, 1973), and thus a higher dispersion indicates higher demand. Extending this stream of research, recent studies use online product review data to examine the relationship between sales and additional metrics of online reviews, such as *valence*, measured as the average rating per product; and *density*, measured as number of reviews per product divided by the number of transactions per product (Dellarocas et al, 2004).

Though researchers agree that online reviews affect sales, there is some disagreement regarding which metrics of online reviews consumers use, how online reviews affect sales, and to what degree. Table 1 presents some of the conflicting results found in the literature regarding the various metrics of online word-of-mouth.

One of the metrics that has yielded mixed results in the literature is volume, measured as the number of reviews. Godes and Mayzlin (2004) studied Usenet conversations about television shows and their relation to Nielsen (viewership) ratings; they found that the dispersion of conversations across different newsgroups has significant explanatory power, but their volume doesn’t. In the context of Yahoo! Movies message board discussions, Liu (2006) found that volume, but not valence, of online conversations has explanatory power on motion picture box office revenues. In the same context, Duan et al. (2005) reached a similar conclusion. Examining book sales, Li and Hitt (2006) found that volume has a positive significant impact on sales in the cross-sectional sales model but a negative significant impact on sales in the fixed-effects sales model.

### Table 1. Examples of Mixed Results for the Relationships between Demand and Metrics of Online Word-of-Mouth: Valence and Volume

<table>
<thead>
<tr>
<th>Metric</th>
<th>Significant</th>
<th>Not Significant /Conflicting Results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Volume</strong></td>
<td>Movie Revenue: Volume has impact on motion picture revenues (Duan et al. 2005; Liu 2006) Movie Revenue: Volume is a proxy of early sales, (Dellarocas et al. 2005) Book sales: An increase in relative volume of reviews at Amazon.com versus BN.com is associated with greater relative sales of a given book at Amazon over time. (Chevalier and Mayzlin, 2006) Beer sales: Volume of reviews for a beer is a good predictor for sales (Clemons et al. 2006)</td>
<td>TV Show Ratings: Volume on discussion forums has no impact on Nielsen Ratings, (Godes and Mayzlin 2004) Book sales: Volume has a positive and significant impact on sales in the cross-sectional sales model but a negative and significant impact on sales in the fixed-effects sales model (Li and Hitt, 2006)</td>
</tr>
</tbody>
</table>
In this paper, we examine the relationship between volume of online reviews and sales. We develop a conceptual model of an online retailer that sells competing products; we assume consumers use volume and valence of reviews to make their purchasing decisions, and derive the demand functions. Based on our conceptual model we identify the following research questions:

1. Does valence of competing products have a cross effect on demand? I.e., does the demand for a product decrease as the valence of a competing product increases?
2. Does the effect of volume on sales differ dependent upon whether the valence (average rating) is perceived to be “good” or “bad” in relation to an anchor point?
3. In the presence of competing products, does the effect of volume on sales depends on the amount of variation between the valence of the product under consideration and the valence(s) of the competing product(s)?

The answers to the above research questions can provide insights regarding the mixed results in the existing literature, and explain why some studies find volume to be significantly related to sales, while others do not find such a relationship.

Using data collected from a large online retailer over a five year period we address the above research questions. We find that valence of competing products has a significant negative effect on sales of the product under consideration. We also find that volume has a positive significant effect on sales of products whose valence is perceived positively by consumers (valence higher than an anchor point), while volume has a negative significant effect on sales of products whose valence is lower than an anchor point. This suggests that when analyzing the effect of reviews on sales, and specifically the relationship between volume of reviews and sales, researchers need to separate the two cases and run two separate models, one with products with high valence and another with products with low valence. Pooling all products together (products with high valence and products with low valence) in one regression model may not capture the relationship between volume and sales, and the two opposing effects might cancel each other out and give the impression that volume does not significantly impact sales.

Finally, addressing our last research question, we show that the use of volume and valence are substitutes when evaluating online review data. When there is enough dissimilarity among the valence scores of competing products, the effect of volume on sales is insignificant. But when the variation among valence scores is low, volume becomes significant. This finding suggests that when the valence metric does not provide sufficient amount of information to differentiate competing products, consumers are more likely to use the volume metric in making their purchase decision.

The rest of the paper is organized as follows. In the next section we present our conceptual model. In Sections three and four we discuss our hypotheses and describe our data set and methodology. In Section five we discuss our empirical results. Finally, in Section six we discuss the theoretical and managerial implications of this research as well as avenues for future research.

**Conceptual Model**

Most online product reviews systems are supported by retail websites or by third parties. On the retailer’s website, consumers can often find more than one product that fits their needs or desired specifications. Previous research shows that sales are associated with positive reviews, and examines the relationship between various summary statistics of the reviews and sales, but does not model specifically the presence of substitutes (competing products). In what follows we model and discuss how reviews, which serve as signals regarding products’ quality, affect the demand for imperfect substitutes when consumers use volume and valence of product reviews to make their purchasing decision.

---

1 The purpose of the conceptual model is to give the background for the Hypotheses, and to explain the use of relative-volume, anchor and substitute-valence in the empirical study. We do not use the demand functions developed in the conceptual model to estimate demand, and we do no use these demand functions in the regression.
Consider a retailer selling two products. The products are imperfect substitutes and thus, when there are no product-reviews, the demand is given by the following general linear demand functions:

\[ D_1(p_1, p_2) = a_1 - bp_1 + dp_2, \quad D_2(p_1, p_2) = a_2 - bp_2 + dp_1. \]  

(1)

According to Equation 1, the demand for product \( i \) decreases with its own price, but increases with the price of the substitute.

Next, we consider how online reviews affect the demand functions. \( R_1 \) denotes the review information for product 1 and \( R_2 \) denotes the review information for product 2. Table 2 summarizes notation used in the paper.

<table>
<thead>
<tr>
<th>Table 2. Notation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( V_i )</td>
</tr>
<tr>
<td>( w_i )</td>
</tr>
<tr>
<td>( p_i )</td>
</tr>
<tr>
<td>( R_i )</td>
</tr>
<tr>
<td>( A_i )</td>
</tr>
<tr>
<td>( E_A )</td>
</tr>
<tr>
<td>( \pi )</td>
</tr>
</tbody>
</table>

In accordance with previous research, we model the effect of reviews on demand using an additive function, \( F \). That is, when product reviews information is available, the demand functions are given by:

\[ D_1(R_1, R_2) = a - bp_1 + dp_2 + F(R_1, R_2) \]
\[ D_2(R_1, R_2) = a - bp_2 + dp_1 + F(R_2, R_1) \]  

(2)

On most retail websites it is easy to compare the number of reviews and the average ratings (valence) across substitutes, but it takes more effort to actually read the individual reviews. Hence, it is reasonable to assume that many consumers use summary statistics, such as the average rating per product, to make their purchasing decision. Accordingly, we assume that \( F \) is a function of the displayed summary statistics of \( R_1 \) and \( R_2 \): Volume and Valence.

**Volume**

The information conveyed by the average rating (i.e. the valence) is perceived to be more reliable if it is based on a larger number of reviews. Since “larger” is a relative term, we normalize number of reviews by taking the relative number of reviews a product received:

\[ w_i = V_i / (V_i + V_j) \]  

(3)

**Valence**

We assume that consumers use an anchor point to determine whether the average rating of a product is “high” or “low”; this assumption is supported by the pricing literature, which shows that consumers establish an anchor point against which to determine whether a price is considered “high” or “low” (Monroe, 1973). We denote the anchor point used by consumers by \( E_A \), and assume it is identical for both products. That is, before observing the reviews/ratings, consumers don’t expect one product to have higher quality than the other product. This assumption is consistent with Dellarocas (2004), in which in a two–period duopoly game, in the first period, before reviews are posted, consumers a-priori believe that the difference between the two qualities is zero.

2 Thus product-reviews can change market shares but do not affect price elasticity, i.e., a marginal change in price would have the same effect on demand with or without the reviews.
The realized value of $A_i$, (average rating for product $i$) can be smaller or larger then $E_A$. If $A_i > E_A$, then $R_i$ should have a positive impact on the demand for product $i$, and the positive effect should be increasing in $w_i$ (the reliability of the result). If $A_i < E_A$, then $R_i$ should have a negative effect on the demand for product $i$, and the magnitude of the effect should be increasing in $w_i$. Thus, the effect of $R_i$ on demand should be proportional to

$$w_i(A_i - E_A). \tag{4}$$

We multiply the deviation by the relative volume$^3$.

**Demand Functions**

Based on the previous analysis, the demand functions are given by Equation 5:

$$D_1(R_1, R_2) = a - bp_1 + dp_2 + c\omega_1(A_1 - E_A) - \beta w_2(A_2 - E_A)$$

$$D_2(R_1, R_2) = a - bp_2 + dp_1 + c\omega_2(A_2 - E_A) - \beta w_1(A_1 - E_A) \tag{5}$$

where $\alpha \geq \beta \geq 0$.

The first three terms in the expression for $D_i$ give the demand for the product in a market with no feedback system. We assume that $b > d$, so that a product’s own price effect is stronger than the cross price effect (if the two prices go up by the same amount, the demand for both products goes down). The last two terms in the expression for $D_i$ capture the effect of the difference in the products’ ratings. If the value of the summary statistic $A_i$ is larger than its anchor value then reviews for product $i$ increase sales of product $i$, but can decrease sales of product $j$. If the value of $A_i$ is smaller than its expected value, then reviews for product $i$ decrease sales of product $i$, but can increase sales of product $j$. To summarize, our conceptual model suggests that:

a. An increase in the valence of a product has a positive effect on its sales while an increase in valences of competing products has a negative effect on the product’s sales.

b. An increase in the number of reviews a product receives can increase or decrease its sales depending on whether the product’s average rating is higher or lower than an anchor value. A deviation that increases the perceived quality of the product will lead to volume coefficient being significant and positive. A deviation that decreases the perceived quality of the product – will lead to volume coefficient being significant and negative.

In what follows, we use a modified version of the classic regression model for the relationship between online reviews and sales, to test (a) and (b).

**Hypotheses**

**Hypothesis 1**

We propose that the way in which consumers use the information about the number of reviews depend on whether the overall sentiment about a product, as exhibited by the average rating, is positive or negative. When valence is high (relative to an anchor point) an increase in the number of reviews should have a positive effect on sales: holding the “high” valence fixed, a higher volume means that more consumers had a favorable opinion about the product. When valence is low (relative to an anchor point), an increase in volume should have a negative effect on sales: holding the “low” valence fixed, a higher volume means that more consumers had a negative opinion of the product. Our conceptual model that drives the rationale behind this hypothesis is consistent with the theory of Bayesian belief updating, according to which consumers have a prior belief about the quality of a given product (Neelamegham and Chintagunta 1999). Consumer’s prior belief is equivalent to the anchor or average expected

---

$^3$ It is important to note that if we were to model the effect of $R_i$ on demand as being proportional to $(w_i A_i - E_A)$ then an increase in $V_i$ (which leads to an increase in $w_i$) would lead to an increase in the demand for product $i$ when $A_i < E_A$, which is not consistent with the fact that a more reliable low average (low relative to the anchor point) should decrease demand.
quality that we introduce in the conceptual model. Consumers update their prior beliefs after they read online reviews, and the degree of updating is proportional to the deviation of the “signal” from the prior belief. In addition, the degree of updating is proportional to the perceived precision of the “signal;” the precision of the online reviews signal is proposed to be proportional to the volume of reviews. Hypothesis 1 follows:

Hypothesis 1: The sign of the effect of volume on demand depends on the sign of the deviation of the valence metric from an anchor point.

Hypothesis 1a: Deviations that increase the perceived quality of the product will be associated with a positive impact of volume.\(^4\)

Hypothesis 1b: Deviations that decrease the perceived quality of the product will be associated with a negative impact of volume.

Hypothesis 2

Most models of competition include a cross effect of prices, that is, the demand for a product decreases with its own price, but increases with the price of competing products. The more similar the products are (i.e., as substitutability increases) the stronger the cross effect. Accordingly, differences in prices can transfer market share from one product to the other.

We propose that online product reviews have a similar cross effect on sales of competing products. The demand for a product increases with its perceived quality, but decreases with the perceived quality of competing products. Hypothesis 2 follows.

Hypothesis 2: The demand for a product increases with its valence but decreases with the average valence of competing products.

Hypothesis 3

Consumers use different decision making strategies in different situations and environments (Payne 1982); this is done in an effort to reduce the cognitive costs associated with decision making (Shugan 1980). Accordingly, consumers are usually willing to settle for imperfect accuracy of their decisions in order to reduce effort (Bettman et al. 1990; Johnson and Payne 1985). Because of this trade-off between effort and accuracy, decision makers frequently choose options that are satisfactory but would be suboptimal if decision costs were zero. This is particularly common when alternatives are numerous and/or difficult to compare (Payne et al. 1993). Behavioral decision theory suggests that decision makers may be inclined to focus more on reducing cognitive effort than on improving decision accuracy (Einhorn and Hogarth 1978; Kleinnuntz and Schkade 1993). Aligned with these prior decision theory findings, we expect the consumers attempt to decrease cognitive effort when assessing online product review information. Consumers prefer situations in which the values of the valences differ from each other, as it allows them to more easily differentiate between products—thus reducing cognitive effort.

We define the valence differential of a product as the absolute value of the difference between the product’s valence and the average valence of the other products in the same product sub-category. That is, if there are \(n\) products in the product sub-category to which product \(i\) belongs, the valence differential of product \(i\) is given by:

\[
A_i = \frac{1}{n-1} \sum_{j \neq i} A_j
\]  

According to the previous discussion, we expect that if a product’s valence differential decreases below a threshold value, the effect of reviews volume on the product’s sales will become significant. On the other hand, we expect that when the valence differential of a product increases above a threshold value, the effect of volume will become insignificant. In other words, consumers substitute volume and valence when evaluating online review data: if the

---

4 Hypothesis 1a and 1b hold whether we use volume or relative volume.

5 We consider products in the same product sub-category to be competing products.
valence metric does not provide sufficient amount of information to differentiate competing products, consumers are more likely to use the volume metric.

For the sake of this study, we calculate the threshold as the average of the valence differentials in a given product sub-category. That is, $T$ is the average of the absolute differences between a product’s valence and the average valence of the other products in the same product sub-category, as given in Equation (7).

$$T = \frac{1}{n-1} \sum_{j \neq i} |A_i - \frac{1}{n} \sum_{j \neq i} A_j|$$  \hspace{1cm} (7)

Hypothesis 3 follows:

Hypothesis 3: The effect of volume on demand depends on the difference between the valence of a product and the average valence of other products in the product sub-category.

Hypothesis 3a: When the difference between a product’s valence and the average valence of the other products in the same category is below the threshold value given in Equation 7, the volume is significant.

Hypothesis 3b: When the difference between a product’s valence and the average valence of the other products in the same category is above the threshold value given in Equation 7, the product’s volume is insignificant.

Data Set and Methodology

Our data for this study was collected from a large online retailer and consists of individual product characteristics, transactions records and user review data from April 16th, 1999 through March 3rd 2005. For this paper, we focus on data related only to electronic products. There are 6 categories within the electronics department, and within each category there are several sub-categories, as listed in Table 3.

<table>
<thead>
<tr>
<th>Product Category</th>
<th>Sub-Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audio &amp; Video</td>
<td>Accessories, Car Audio &amp; Video, DVD &amp; VCR, Home Theater, MP3 &amp; iPod Accessories, MP3 Players &amp; iPods, Portable CD &amp; DVD, Radios &amp; Clock Radios, Speakers &amp; Subwoofers, Stereos &amp; Equipment, Televisions</td>
</tr>
<tr>
<td>Cameras &amp; Optics</td>
<td>Accessories, Binoculars &amp; Optics, Camcorders, Digital Cameras, Film Cameras, Memory Cards &amp; Readers, Surveillance</td>
</tr>
<tr>
<td>Printers &amp; Scanners</td>
<td>Printers, Printer Accessories, Scanners</td>
</tr>
<tr>
<td>Monitors</td>
<td>LCD Flat Panel Monitors, CRT Monitors, Monitor Accessories</td>
</tr>
<tr>
<td>Telephones</td>
<td>Cell Phones, Cell Phone Accessories, Cordless Phones</td>
</tr>
</tbody>
</table>
For each transaction in our sample, we collected the following information: product’s ID and sub-category, the price charged for each product, the shipping charges and taxes, and the date and timestamp of the transaction. In addition, for each product available for sale on the retailer’s website, we collected the number of reviews, the rating score associated with each review (on a scale of 1 to 5, with 5 being the best), the text of the review, and the year in which the review was posted\textsuperscript{6} to the website.

In addition to the raw data, i.e. the individual reviews, we calculated several aggregate measures of the reviews associated with each product. These measures include: Valence, Volume, Relative Volume, Substitute Valence, and Valence Differential

**Valence:** In order to calculate the valence for each product we take the average of all of the rating scores from all the reviews submitted for that product up to a given date. For example, to find the valence of a product in 2002 we use the reviews posted for the product from April 1999 up to the end of 2002; to find the valence of the same product in 2003 we use the reviews posted from April 1999 up to the end of 2003. We can calculate the valence only on a yearly basis, as we know only the posting year for each review (there is no data for month and day).

Table 4 shows the average valence (across all the products in the electronics category) and the average variance in ratings per product for the different years in our data set.

| Table 4. The Average Valence and the Variance in Ratings per Product by end of Each Year |
|---------------------------------------------|---|---|---|---|---|---|---|
|                                           | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 |
| Average Valence                           | 4.36 | 4.37 | 4.37 | 4.42 | 4.46 | 4.48 | 4.47 |
| Variance                                  | 0.9628 | 1.0669 | 1.093 | 1.073 | 1.06 | 1.053 | 1.045 |

It is interesting to note that contrary to findings of prior literature regarding book reviews (Li and Hitt, 2004), which showed that average rating decreases over time, the average rating slightly increases over time for electronic products.

**Volume and Relative Volume:** Aligned with the conceptual model, we investigate both absolute volume (i.e. number of reviews submitted for a product) and relative volume. We calculate the relative volume for a given product by dividing the total number of reviews posted for the product by the total number of reviews for all of the products in the same product sub-category. As with valence, we can calculate the volume of a product for each of the years in our data set. For Example, to find the reviews volume for a product in 2002 we use the reviews posted from April 1999 up to the end of 2002; to find the reviews volume for the same product in 2003 we use the reviews posted from April 1999 up to the end of 2003.

Tables 5 and 6 below show the average of the number of reviews per product, the maximum number of reviews a produce accumulated, and the standard deviation in the number of reviews submitted for products in the electronic category in each of the years in our data set. Table 5 presents information for products with high valence, i.e. valence that is higher than 3, and Table 6 presents information for products with low valence, that is products with valence lower than 3.

| Table 5. Cumulative Volume of Reviews for Products with High Valence |
|---------------------------------------------|---|---|---|---|---|---|---|
|                                           | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 |
| Average # of Reviews per Product           | 1.023 | 1.146 | 1.266 | 1.595 | 2.340 | 3.842 | 5.439 |
| Standard Deviation                         | 0.149 | 0.402 | 0.604 | 1.085 | 2.318 | 4.156 | 6.201 |
| Max # of Reviews                           | 2    | 4    | 9    | 16   | 24   | 34   | 39   |

\textsuperscript{6} Unfortunately, the retailer does not time stamp the reviews, rather, they mark that someone approved the review with the year in which it was done.

\textsuperscript{8} Twenty Eighth International Conference on Information Systems, Montreal 2007
Table 6. Cumulative Volume of Reviews for Products with Low Valence

<table>
<thead>
<tr>
<th></th>
<th>1999</th>
<th>2000</th>
<th>2001</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average # of Reviews per Product</td>
<td>1.000</td>
<td>1.057</td>
<td>1.102</td>
<td>1.227</td>
<td>1.455</td>
<td>1.620</td>
<td>1.804</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.000</td>
<td>0.254</td>
<td>0.388</td>
<td>0.659</td>
<td>1.238</td>
<td>1.725</td>
<td>2.220</td>
</tr>
<tr>
<td>Max # of Reviews</td>
<td>1</td>
<td>3</td>
<td>7</td>
<td>10</td>
<td>12</td>
<td>19</td>
<td>21</td>
</tr>
</tbody>
</table>

Comparing the two tables above we see that the average number of reviews for products with high valence is consistently higher than the average number of reviews for products with low valence. In addition the number of reviews for products with high valence increases at a faster rate than the number of reviews for products with low valence. This is clearly exhibited in Figure 1. The same holds when comparing the standard deviation in volume, as is exhibited in Figure 2: there is higher standard deviation in the volume of products with high valence than in the volume of products with low valence, and the standard deviation increases over time for both groups, but at a faster rate for the first one.

Figure 1. Average Cumulative Volume; Products with Valence Above 3 vs. Products with Valence Below 3

Figure 2. Standard Deviation in Cumulative Volume; Products with Valence Above 3 vs. Products with Valence Below 3
Substitute Valence: aligned with the conceptual model, we investigate the valence of the substitute products by taking the average rating of all other products that are in the same product sub-category, and are therefore considered substitutes for the given product.

Valence Differential: We calculate the valence differential as the difference between a product’s valence and the average valence of other products in the same product sub-category.

For the three hypotheses, we run an aggregate model where we use one year of sales data from April 2004 to March 2005 along with review data posted between 16 of April 1999 and 3-March 20047, as is exhibited in Figure 3.

![Figure 3: Data Used in the Regression Model](image)

We incorporate a time lag between the last date of review data and the start date of sales data in order to separate the impact of reviews on sales. When reviews and sales are examined in a cross-sectional analysis, the question of endogeneity arises; thus one may ask whether a product with more sales also has a higher volume of reviews, such that sales volume is driving review volume, rather than review volume impacting sales volume. By only including reviews from the period preceding the transaction period, we can isolate the impact of reviews on sales: because the sales are occurring after the reviews, it is not possible that sales volume is driving review volume.

The empirical model specifications are extended from prior literature on online reviews and sales (e.g. Chevalier and Mayzlin, 2006). The extensions are guided by the conceptual model:

1. We include the average valence of substitutes as an independent variable. We do not include relative volume of the substitutes due to multicollinearity (it equals 1 minus the relative volume of the product)
2. The conceptual model represents the relationship between relative volume and perceived valence as multiplicative. In order to empirically examine this multiplicative relationship using a linear model, we take the log-log model.
3. In order to separate out the effect of valence scores that are above the anchor point from valence scores that are below the anchor point, we utilize a dummy variable (AnchorDummy), which equals 0 if Valence<Anchor and equals 1 otherwise (where Anchor is the anchor point for evaluating the valence).

Combining 1 though 3 we arrive at the following empirical model:

\[
\text{Ln}(Sales_i) = a - \alpha_0 \text{Ln}(p_i) + \alpha_1 \text{Ln}(\text{relativeVolume}_i) + \alpha_2 \text{Ln}(\text{Valence}_i - \text{Anchor}) \\
- \alpha_3 \text{Ln}(\text{SubsValence}_i - \text{Anchor}) + \alpha_4 \text{AnchorDummy}
\]

7 Due to a change in the online review storage, the only reviews that are associated with a day rather than a year are those of March 3, 2005. Since the reviews are in chronological order, we are able to use these data as a cut-off point for the aggregate model.
Results

Hypothesis 1 argues that the effect of volume on demand depends on whether the product’s valence is perceived in a positive or negative way by consumers. To assess Hypothesis 1, we create a dummy variable which separates products that have a valence score above the anchor point from products that have a valence score below the anchor point. For the sake of this study, we use the midpoint of the 5-point rating scale, the number 3, as the anchor point. Table 7 shows our empirical results which support Hypothesis 1. The effect of volume on demand is significant and positive in the case where valence is greater than the anchor point; whereas the effect of volume is significant and negative in the case where valence is less than the anchor point. Thus an increase in the number of ratings is associated with an increase in sales when the average rating is above the anchor point. On the other hand, an increase in the number of ratings is associated with a decrease in sales when the average rating is below the anchor point. In addition, supporting Hypothesis 2, the effect of the deviation of the valence from the anchor is positive and significant in both cases, which is expected as an increase in valence should be associated with an increase in the perceived quality of the product, whether the valence is low or high. In addition we see from Table 7 that the average valence across the substitutes (other products in the same sub-category) has a negative and significant effect on sales. That is an increase in the valence of substitutes decrease sales.

| Table 7. Empirical results for Hypothesis 1 |
|----------------------------------------|-----------------|-----------------|-----------------|-----------------|
|                                      | Low valence case | High valence cases |
|                                      | (Valence smaller than anchor) | (Valence larger than anchor) |
| Coef. | P>t            | Coef. | P>t            | Coef. | P>t            |
| Price | -0.70**        | 0.004 | -0.85*         | 0.000 |
| Valence - Anchor                     | 0.20**         | 0.037 | 0.18***        | 0.003 |
| Relative Volume                      | -0.33**        | 0.021 | 0.24**         | 0.036 |
| Substitute_Valence-Anchor            | -0.62**        | 0.032 | -0.76**        | 0.028 |
| N                                  | 569            | 4388  |                | 0.4364         |
| Adj Rsq.                            | 0.3536         |       |                |                |

Hypothesis 3 states that the volume is significant if the difference between the valence of the product and the average valence of the other products in the same product sub-category is below a threshold value, and is insignificant if the difference between the valence of the product and the average valence of the other products in the category is above the threshold value.

We take the results of hypothesis 1 into account, and run the analysis separately for electronic products that have an average rating lower than the anchor of 3 and products that have an average rating higher than the anchor.

As Table 8 shows, the volume of ratings is not significant in the case where the difference between the valence of the product and the average valence of the other products in the product sub-category is above the mean (the threshold given in Equation 7 is the mean of the differences between a product’s valence and the average valence of the competing products). Thus, volume is not significant when the information conveyed by the valence scores differentiates products enough for consumers to use just the valence to make a purchase decision. Confirming the second part of hypothesis 3, for products for which the difference between the valence and the average valence of the other products in the sub-category is below the mean, the volume is significant. Thus, when the valence does not convey enough information to differentiate products, the volume becomes a significant factor in consumers’ purchase decisions. Re-confirming Hypothesis 1, for the low-valence case (products with valence below the anchor value of 3) volume is negative and significant while for the high-valence case (products with valence above 3) volume is positive and significant.
Table 8. Effect of Volume on Sales for Products with Low Valence Differential and for Products with High Valence Differential

<table>
<thead>
<tr>
<th></th>
<th>Low Valence Differential</th>
<th></th>
<th>High Valence Differential</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low Valence Products</td>
<td>High Valence Products</td>
<td>Low Valence Products</td>
<td>High Valence Products</td>
</tr>
<tr>
<td>Coef. P&gt;t</td>
<td>-0.826*** 0.000</td>
<td>-0.865*** 0.000</td>
<td>-0.863*** 0.000</td>
<td>-0.851*** 0.000</td>
</tr>
<tr>
<td>Price</td>
<td>-0.038*** 0.000</td>
<td>0.062** 0.046</td>
<td>-0.106 0.112</td>
<td>0.023 0.214</td>
</tr>
<tr>
<td>Relative Volume</td>
<td>0.067** 0.0008</td>
<td>0.044* 0.053</td>
<td>0.072** 0.045</td>
<td>0.050** 0.015</td>
</tr>
<tr>
<td>Valence-Anchor</td>
<td>339 0.7943</td>
<td>2434 9010</td>
<td>227 0.8565</td>
<td>1951 0.8513</td>
</tr>
</tbody>
</table>

Note that we do not include the substitute valence in this set of regressions, as it is already taken into account through the use of the dummy variable which splits the sample according to whether the difference between the valence of the product and the average valence of the substitute products is above or below the threshold.

To examine the robustness of these results we repeat the specification used in the analysis of each of the hypotheses, but with a modified anchor dummy variable. We utilized an anchor of 4 rather than 3, such that valence scores above 4 were considered “positive”, and the valence scores below 4 were considered “negative”; in addition, we tried an anchor point of 2, such that the valence scores above 2 were considered “positive” and the valence scores below 2 were considered “negative”. The results generally hold using these alternative anchors. Lastly, for the results presented in this paper we ran the models using only data for products that had reviews. To test the robustness of the results, however, we also ran a specification which included a “no review” dummy variable. Using the “no review” dummy makes no difference in the high valence case, as by definition, if the score is above 3, there are reviews. Thus all of the results are the same in the high valence cases. In the low valence case, the no review dummy is significant, and the remaining results are consistent with those shown in the paper.

Discussion

Online word-of-mouth is used in many electronic commerce sites so that consumers can inform other consumers regarding products quality and characteristics. Prior research has established that online word-of-mouth influences consumer behavior (Chevalier and Mayzlin 2006; Senecal and Nantel 2004) and future sales (Godes and Mayzlin 2004; Duan et al. 2005; Liu 2006). However, the literature has conflicting results as to whether volume of online word-of-mouth has a significant impact on consumers’ purchasing decisions (Mayzlin 2004; Duan et al. 2005; Liu 2006) and whether the impact is positive or negative. This paper improves our understanding of the relationship between volume of online product reviews and sales. Specifically, our study sheds light on the following three aspects: 1) The sign of the relationship between a product’s review volume and its sales, and its dependency on the product’s valence; 2) the cross effect of valence scores; and lastly, 3) the relationship between volume and valence, and whether consumers substitute between the use of the two metrics.

We estimate the models using metrics obtained from user reviews posted on the website of a large online retailer, together with actual transaction information obtained from the retailer. Focusing on the electronics product category, we show that the impact of volume on product sales is affected by two different factors: 1) whether the valence of the product’s rating is above or below an anchor point; and 2) the variation between a product’s valence and the average valence of the other products in the product category.
First, we show that volume has a significant negative impact on sales when valence is below an anchor point, whereas it has a significant positive impact when valence is above an anchor point. This observation may shed light on the conflicting results regarding the effect of volume on sales reported in the literature. Perhaps in situations where volume was found to be insignificant the valence scores were so variant that the cases in which volume had a significant positive impact were being cancelled out by the cases in which volume had a significant negative impact. Furthermore, this has potentially important implications related to products that consumers rate negatively, as allowing additional consumers to post negative reviews may result in a decrease in sales for those products.

Second, we show that the volume has statistical significance in the case where the difference between the valence of a product and the average valence of other similar products is low. We expect that in most product categories, the difference between the valence across products is not very high, in fact even when we do not control for such variance we see that volume is significant (table 7). However, when the difference of the valence across products is high, then consumers use the valences to make the purchase decision, and volume becomes insignificant.

It was the goal of our work to help reconcile some of the inconsistencies among previous studies with respect to when volume of online word-of-mouth is statistically significant with respect to online consumer behavior. Specifically, we show that, volume has different impacts based on the value of the valence score and the variation across valence scores of competing products. We conclude by pointing out some opportunities for future research. First, while one of the strengths of this paper is that we do take into account competing products sold on the same website, we are not able to incorporate the impact of competition from other online retailers; such would be a good avenue for future research. Second, motivated by the fact that in this paper we shows some substitutability between volume and valence, in future work we plan to include additional metrics in our analytical model, and examine how consumers choose among the different metrics they might use to make the purchasing decision. Third, although the current study only looks at electronic products, the findings may likely generalize to other products as well. It would, thus, be interesting to investigate to what extent our models are applicable in the context of other classes of goods such as books, music, and home goods.

Conclusion

Consumer use of online ratings data is an important and promising area of research. The literature that examines online ratings in relation to online sales reports mixed results regarding the significance of volume towards online purchasing behavior. In this paper, we present important findings for academics and managers alike. Namely, we suggest that the impact of online review volume cannot be examined in isolation. Rather, the relationship between a product’s reviews volume and sales depends on various factors, including: the product’s valence score, volume level, and the differential between the product’s valence score and valence scores of substitute products. These results offer an important first step into understanding the interdependence of various metrics of online ratings in relation to online sales.

References


