Why do Online Product Reviews have a J-shaped Distribution? Overcoming Biases in Online Word-of-Mouth Communication[1]

Nan Hu

School of Information Systems

Singapore Management University

hunan@smu.edu.sg

Paul A. Pavlou*

Anderson Graduate School of Management

University of California, Riverside

900 University Ave, Anderson Hall 226

Riverside CA 91251

Phone: (213) 268-2259

Fax: (951) 827-9370

paul.pavlou@ucr.edu

Jennifer Zhang

College of Business Administration

University of Toledo

2801 W. Bancroft St. MS#103

Toledo, OH 43615

Phone: (419) 530-2245

Fax: (419) 530-2290

Jennifer.zhang@utoledo.edu

A Submission to *Marketing Science*

February 2007

* Corresponding Author

Why do Online Product Reviews have a J-shaped Distribution? Overcoming Biases in Online Word-of-Mouth Communication

ABSTRACT

Online word-of-mouth communication in the form of product reviews is a major information source for consumers and marketers about product quality. The literature has used the mean of online reviews to predict product sales, assuming that the mean reflects product quality. However, using a combination of econometric, experimental, and analytical results, we show that the mean is a biased estimator of product quality due to two self-selection biases (purchasing and under-reporting bias). First, econometric results with secondary data from Amazon.com show that almost all products have an asymmetric bimodal (J-shaped) distribution with more positive than negative reviews. Second, experimental results where all respondents wrote reviews show that their reviews have an approximately normal distribution with roughly equal number of positive and negative reviews. This implies two biases: (1) *purchasing bias* - only consumers with favorable disposition towards a product purchase the product and have the opportunity to write a product review, and (2) *under-reporting bias* - consumers with polarized (either positive or negative) reviews are more likely to report their reviews than consumers with moderate reviews. This results in a *J-shaped* distribution of online product reviews that renders the mean a biased estimator of product quality. Third, we develop an analytical model to derive the conditions for the mean to become an unbiased estimator of product quality. Based on these conditions, we build a new model that integrates three distributional parameters - *mean*, *standard deviation*, and the *two modes* of the online product reviews (to overcome under-reporting bias) and *product price* (to overcome purchasing bias). This model is shown to be a superior predictive model of future product sales compared to competing models.

Key Words: Product Quality, Word-of-Mouth Communication, Online Product Reviews, Bi-Modal Distribution, Unimodal Distribution, Online Consumer Behavior, Internet Marketing, E-commerce, Sales Forecasting, Econometric Models, Experimental Economics, Regression and other Statistical Techniques.

INTRODUCTION

Pioneered by Bass' (1969) seminal work, word-of-mouth (WOM) communication is considered a valuable marketing resource for both consumers and marketers. WOM communication is defined as all informal exchange of information among consumers about the characteristics, usage, and ownership of products and services, and it is considered a major driver of consumer adoption and diffusion of new products, particularly for late adopters (e.g, Banerjee 1993, Biyalogorsky *et al*. 2001, Brown and Reingen 1987, Eliashberg *et al*. 2000, Eliashberg and Shugan 1997, Krider and Weinberg 1998, Buttle 1998, Herr *et al*. 1991, Mahajan *et al*. 1984, Sheth 1971).

While WOM has been traditionally spread among acquaintances through personal "contagions", the Internet has dramatically increased the scale of WOM communication (Dellarocas 2003). Online WOM communication in the form of *online product reviews* has become a major informational source since online product reviews can reach far beyond traditional settings and reach a virtually infinite number of consumers. Indeed, the marketing literature suggests that consumers do pay attention to online product reviews and act upon them to make purchasing decisions (Chatterjee 2001, Chevalier and Mayzlin 2006, Senecal and Nantel 2004). Reichheld (2003) even claims that the customer's propensity to recommend a product to others – termed *referral value* - is the most important success measure in marketing today. Chen and Xie (2005) even examined when and how firms adapt their marketing strategies to online product reviews. Therefore, firms like Amazon.com and Circuitcity.com encourage their consumers to read and write product reviews, while other firms, such as Epinions.com and BizRate.com specialize on collecting, synthesizing, and disseminating online product reviews.

While there is an emerging literature on the role of online product reviews on consumer behavior and, consequently, on sales (e.g., Chevalier and Goolsbee 2003, Chevalier and Mayzlin 2005, Dellarocas *et al*. 2004, Duan *et al*. 2005, Godes and Mayzlin 2004, Goldenberg et al. 2001, Li and Hitt 2004, Liu 2006), the results are often inconsistent. While many studies show that online product reviews have a significant effect on product sales (e.g, Chevalier and Goolsbee 2003, Clemons *et al*. 2006, Dellarocas *et al*. 2004, Li and Hitt 2004), other studies (e.g., Chen *et al*. 2004, Duan *et al*. 2005, Liu 2006) show that it is the *volume* of product reviews that significantly influences sales, while the *valence* does not show any significant effects. Chevalier and Mayzlin (2006) examined the effect of online product reviews on book sales on Amazon's and Barnes and Noble's websites, showing that an improvement in a book's mean of online product reviews enhances book sales. Li and Hitt (2004) show similar results with online product reviews and book sales with data from Amazon.com. In contrast, Chen et al. (2004) show that the mean of online product reviews does *not* have a significant impact on future book sales, but only the number of online product reviews does. Dellarocas et al. (2004) use a Bass diffusion model to examine how *user movie reviews* taken from the box office history predict the two Bass parameters (*p*: external influence factor or coefficient of publicity, and *q*: internal influence factor or WOM coefficient). The authors find that the mean of consumer movie reviews is a significant predictor of *q*, but an insignificant predictor of *p*. In contrast, from a similar source of data, Duan et al. (2005) and Liu (2006) find that the mean of consumer movie reviews does *not* adequately predict box office revenue. Finally, Clemons el al. (2006) show that the mean of beer quality correlates with sales growth, but the explanatory power is only based on the top-quartile of the beer quality reviews.

Summarizing the literature, what is common across virtually all studies is that they use the *mean* of the product reviews to predict product sales,[2] assuming that the mean is an unbiased estimator of product quality. However, to the best of our knowledge, this assumption has never been tested, neither empirically or analytically.

Product quality is defined as the collective assessment of a product's perceived value across consumers. Product quality reflects the product's intrinsic value, which is the actual value or inherent worth of the product. Perceived value is determined by many variables, such as taste, mood, social status, among others. Therefore, product quality is in the eye of the beholder (consumer), particularly for products where taste is a major factor, such as books, music, and movies examined in this study. Thus, each consumer's product review is an assessment of her own perceived value, and product quality is thereby the aggregate of all consumer perceived values, independent of their individual heterogeneous tastes. The issue is whether the mean of the online product reviews is an unbiased or a biased estimator of product quality. We formally introduce the first set of research questions:

- **Is the mean of online product reviews a biased estimator of product quality?**

- **If not, what are the sources of this bias, and how can we overcome them?**

To shed light on these two questions, we first empirically test the distribution of online product reviews for many product categories with longitudinal secondary data from Amazon.com (the largest online retail store). Our econometric results reveal that virtually all online product reviews follow an asymmetric bimodal (J-shaped) distribution with more extreme positive (5-stars) than extreme negative (1-star) reviews on a 5-star Likert-type scale. The asymmetric, positively skewed distribution of online product reviews has also been reported by Eliashberg and Shugan (1997), Chevalier and Mayzlin (2006), and Liu (2006), and it has also caught the attention of the business press (Kadet 2007). For these online product reviews, the mean may be a biased estimator of product quality. We offer three explanations for the J-shaped distribution: (1) Consumers have polarized, positively skewed tastes, and the J-shaped distribution reflects the true state of nature; however, this explanation suggests that most products in the market are outstanding; (2) consumers are overconfident (Admati and Pfleiderer 2004) in their product reviews, and the J-shaped distribution reflects consumers exaggerating their overconfident reviews; however, this would result in a symmetric distribution of online product reviews; (3) there are self-selection biases: First, since only consumers with higher product quality valuations are likely to purchase any given product, only consumers who purchased a product have the opportunity to write reviews, which are thus predominantly positive. A similar explanation was offered by Chevalier and Mayzlin (2006). Purchasing bias causes the asymmetry in the distribution of online product reviews, and it also inflates the mean. Second, among consumers who purchased a product, consumers with polarized opinions (either extreme positive, 5-star or extreme negative, 1-star) are more likely to write reviews to "brag-or-moan" than those with moderate (2, 3, or 4-star) reviews (under-reporting bias). This under-reporting bias causes the bimodal distribution, but it does not necessarily cause an asymmetry in the distribution, or influence the mean of the online product reviews in any direction. However, taken together, we argue that these two independent self-selection biases jointly explain the observed bimodal, J-shaped distribution.

---

[2] Notable exceptions are the study of Chevalier and Mayzlin (2006) that uses the percentages of 1-star and 5-star reviews, and the study of Liu (2006) that uses the percentages of positive and negative review messages.

To support our explanation of the J-shaped distribution of online product reviews, we performed a controlled experiment in which all respondents were asked to report their reviews for a randomly-selected product. When all consumers reported their product reviews, the experimental results showed a unimodal (approximately normal) distribution, while Amazon's corresponding distribution for the very same product was J-shaped. Also, the mean of the experiment's product reviews was significantly lower than the mean of Amazon's online product reviews. These results imply that the J-shaped distribution of online product reviews is not due to "polarized tastes" or "overconfidence," but rather due to the fact that consumers who purchase the product are more likely to write a product review (purchasing bias), while consumers with moderate reviews are less motivated to exert the time and effort to report their review (under-reporting bias). This confirms that the mean of online product reviews based on a truncated sample (bimodal, J-shaped distribution) is likely to be a biased estimator of product quality, and potentially a poor predictor of future product sales. These findings lead to our second set of research questions:

- **Under what conditions can the mean of the online product reviews become an *unbiased* estimator of product quality?**
- **Does a model that accounts for the identified biases (purchasing and under-reporting) of the online product reviews better predict future product sales than the simple mean and other competing models?**

To answer these research questions, we build an analytical model to theoretically derive an unbiased estimator of product quality and specify the conditions for the mean of online product reviews to reflect product quality, given the varied probability of consumers voluntarily reporting online product reviews (depending on their reviews, moderate or polarized) and given that consumers with low product valuations are less likely to purchase the product and write product reviews. The analytical model infers three distributional parameters - the *mean*, *standard deviation*, and the *two modes* of the online product reviews to overcome under-reporting bias, and *product price* to overcome purchasing bias. These variables are needed to properly estimate product quality by overcoming the two self-selection biases that give rise to the J-shaped distribution. Based on the analytical results, we propose a new model with the three distributional parameters and product price to predict future product sales, and we compare this model to a set of competing models (simple mean, weighted mean, percentage of 1-star and 5-star reviews). Future product sales are used as a tangible proxy for product quality since consumers are likely to purchase products they perceive to create value to them, and the collective assessment of a product's perceived value across many consumers is likely to predict its market value, actual market demand, and future sales.

The results show that the proposed model with the *mean*, *standard deviation*, the *two modes* of the online product reviews, plus *product price* better predicts future product sales compared to several competing models. We explain these results on the basis that consumers are rational, and they attempt to overcome the two identified self-selection (purchasing and under-reporting) biases when forming their product quality perceptions and making their purchasing decisions. Therefore, we conclude that consumers are "smart," and they do not solely rely on single-point estimators, such as the mean, which are readily available to them, but they try to overcome these two sources of self-reporting biases by utilizing other factors, such as the proposed distributional parameters.

The paper makes four key contributions: First, by demonstrating that the distribution of online product reviews is asymmetrically bimodal (J-shaped), it shows that the mean of online product reviews is a biased estimator of product quality. Second, by experimentally showing that product reviews from all consumers have a normal distribution whose mean is significantly lower than Amazon's reviews, it shows that the J-shaped distribution of online product reviews is primarily due to self-selection (purchasing and under-reporting) biases. Third, it analytically derives an unbiased estimator of product quality by proposing a number of variables needed to help overcome the two-self selection biases. Finally, it proposes a new model based on results of an analytical model to predict future product sales. Taken together, these contributions help identify the biases in the nature of online product reviews and derive models for better estimating product quality and predicting future product sales.

It is important to note that comprehensive product reviews (such as those reported in this study from Amazon) are only available in online marketplaces, and comparable data are seldom available in traditional offline markets. Therefore, it is difficult to examine the nature of *offline* product reviews and identify differences and similarities with *online* product reviews. However, we have no reason to believe that the results from online product reviews would not be generalizable, and our results could probably apply to both online and offline markets. Still, to avoid concerns for generalizability, we restrict our discussion to online product reviews and online consumer behavior.

The paper proceeds as follows: Section 1 describes the nature of online product reviews from secondary data from Amazon, empirically demonstrating that almost all online product reviews follow a J-shaped distribution. Section 2 presents the controlled experiment that suggests that the J-shaped distribution of online product reviews is due to self-selection (purchasing and under-reporting) biases, implying a truncated distribution, and thus a biased estimator of product quality. Section 3 describes the analytical model that derives the distributional parameters for the mean of online product reviews to serve as an unbiased estimator of product quality. It also introduces new model that aims to better predict future product sales, and it compares this model with several competing models with single-point estimators. Finally, we discuss the paper's contributions and implications for theory and practice.

SECTION 1. ECONOMETRIC STUDY

Nature and Distribution of Online Product Reviews

To examine the nature and distribution of online product reviews, a random sample of books, DVDs, and videos was chosen. Product information and their corresponding consumer reviews were collected from Amazon.com between February and July of 2005 using Amazon Web Service (AWS).[3] Table 1 provides the descriptive statistics of our sample, which includes products sold on Amazon.

Table 1: Descriptive Statistics of Amazon's Data

Product Category	Number of Products	Number of Online Product Reviews	Mean of Online Product Reviews
Books	32,878	967,075	4.02
DVDs	17,978	2,034,552	4.19
Videos	28,983	1,248,992	3.99

---

[3] The data analysis is based on data collected prior to 7/15/2005. For robustness, we conducted the same analysis based on longitudinal data collected prior to 2/18, 3/18, 4/15, 6/15 of 2005, which rendered similar results.

Figure 1 shows that 78%, 41%, and 66% of books, DVDs, and videos have less than 20 reviews. To capture the distribution of online product reviews and make a meaningful comparison across different product categories, we first assured that each product has an adequate number of reviews. Therefore, we excluded products with fewer than 20 product reviews. We also tried other cutting points (e.g., 30, 40, and 50), and the results regarding the nature of the underlying (either unimodal or bimodal, J-shaped) distribution are qualitatively the same.

**Figure 1: Cumulative Distribution of Online Product Reviews for Books, DVDs, and Videos**
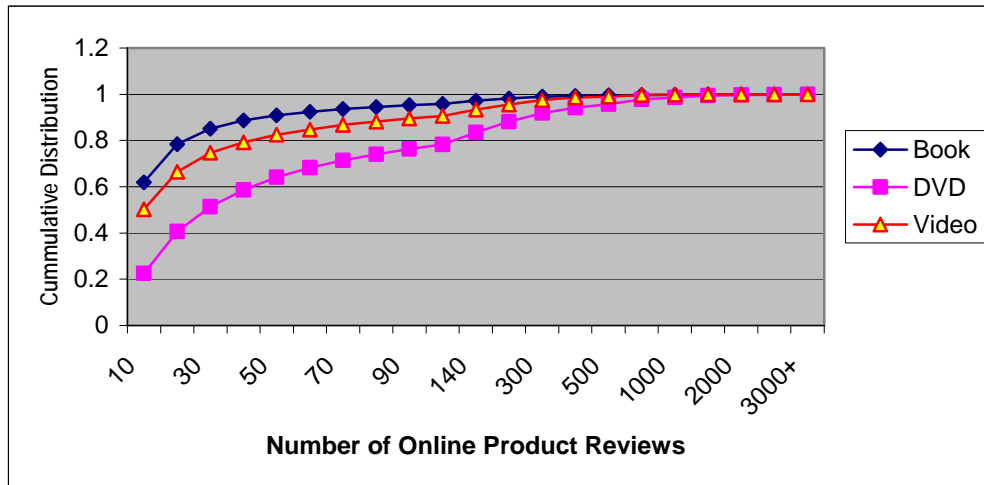
On Amazon, consumers can only report an integer product review on a 1-star to 5-stars Likert-type scale, anchored at 1-star = least satisfied and 5-stars = most satisfied. As shown in Figure 2a, about 78%, 73%, and 72% of the online product reviews for books, DVDs, and videos are greater or equal to 4-stars, respectively. This is true for the whole sample of products, as well as for products in each individual category. These results are consistent with Chevalier and Mayzlin (2006) who show that reviews for books on Amazon and Barnes & Noble tend to be overwhelmingly positive. These results can be explained by the fact that consumers with higher product valuations are more likely to purchase the product, and they are thus more positively disposed to write a positive product review. In contrast, consumers with lower product valuations and negative disposition toward the product are less likely to purchase the product, and they will not have the opportunity to write a negative product review.

Figure 2a: The Distribution of Online Product Reviews for *all* Books, DVDs, and Videos
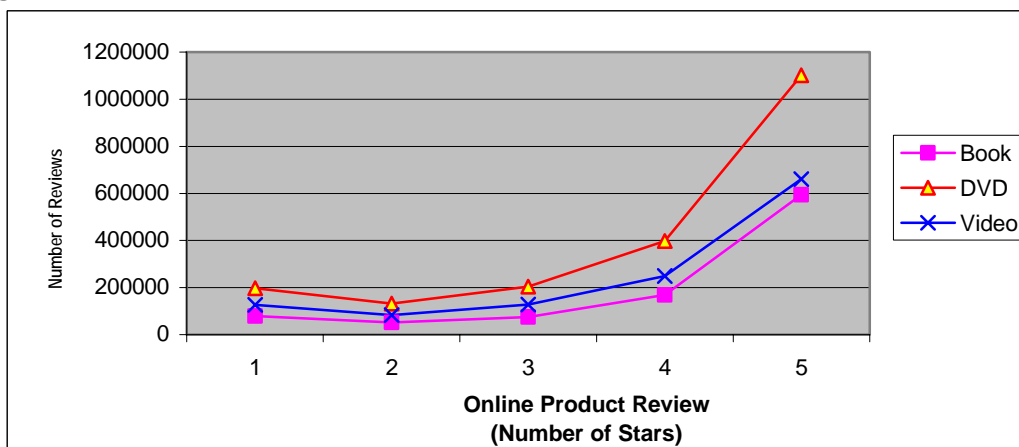
For each product category, Figure 2a shows a bimodal (J-shaped) distribution of online product reviews, which is contrary to Jensen's (1969) claim that any test with "a large number of products" will end up having a *normal* distribution. Figure 2b shows the distributions of three randomly-selected products in each of the three categories with more than 2000 reviews. The results show that these products still have a bimodal distribution, confirming that the observed J-shaped distribution is not due to the small number of reviews.

Figure 2b: The Distribution of Online Product Reviews for Books, DVDs, and Videos with > 2000 Reviews
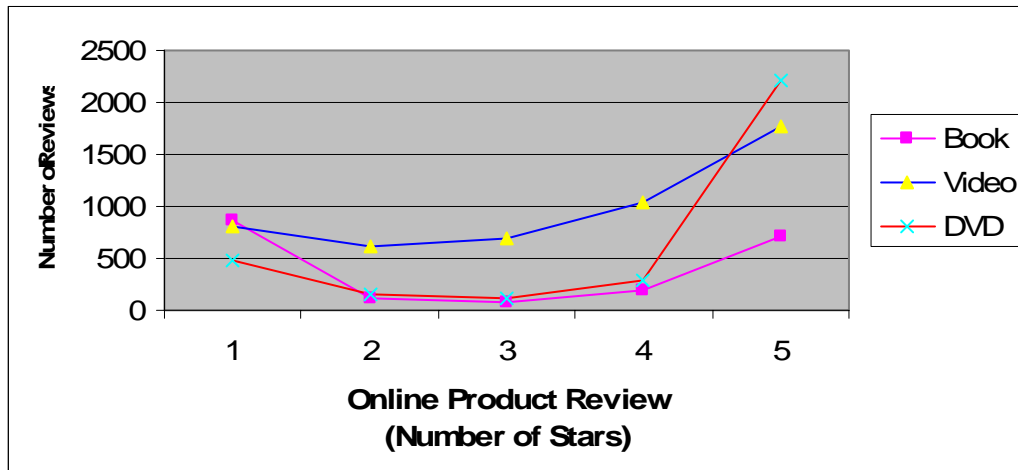
To test the robustness of the J-shaped distribution, we first pulled together all products with an average review of 3-stars within each product category, and we plotted the distribution of product reviews (Figure 3).[4]

Figure 3: The Distribution of Books, DVDs, and Videos with Mean Product Review of 3-Stars
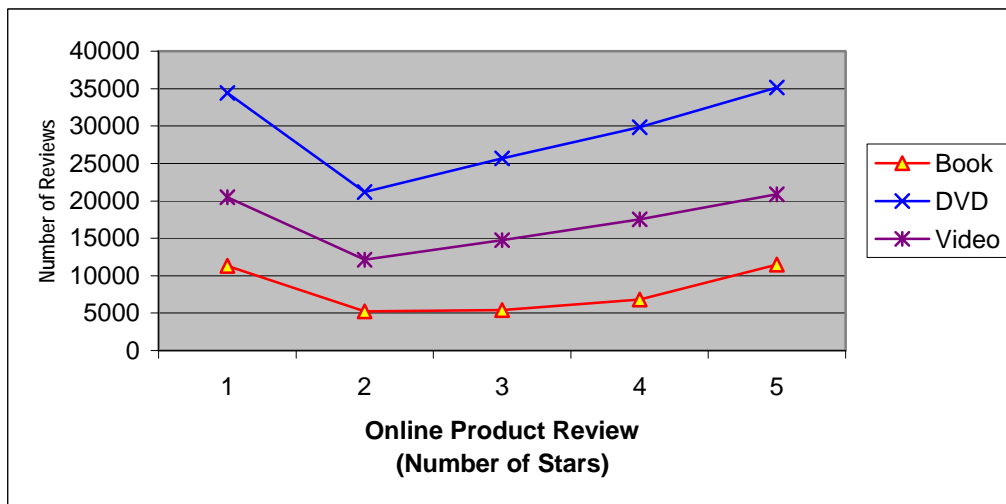

Figure 3 also shows a bimodal, J-shaped distribution of online product reviews for most products. Specifically, for products with a mean of 3 stars, the online product reviews follow a bimodal distribution, which

---

[4] We calculated the mean of each product's online reviews based on all observations. Since the mean can be decimal number, such as 1.2 or 2.1, we used the following classification: If the mean of the online product reviews was between 1 and 1.5, we classified the product into a group with a mean=1; if the mean was between 1.5 and 2, we classified the product into a group with a mean=1.5, etc. We also tried to classify products whose mean was around 1 (e.g., between 0.9 and 1.1), into a group with mean=1, and the results were very similar.

is a strong indication that the distribution is *not* normal. We argue that the bimodal, J-shaped distribution indicates a positively skewed polarization of online product reviews.[5] These results thus suggest that the mean of online product reviews is a biased estimator of product quality since the product reviews do not converge to the mean. The mean rather reflects the balance of diverse product reviews. In other words, when a product's mean is around 3 stars, it does not suggest that consumers generally agree that this is an average product. It rather suggests that many consumers think this is an excellent product, while some think it is a poor product.

Testing for Normality and Bimodality

Having graphically shown that the distribution of Amazon's online product review is not normal, we next use the Kolmogorov-Smirnov, the Cramer-von Mises, and the Anderson-Darling tests to formally verify whether the distribution is normal. Chakravart, Laha, and Roy (1967) proposed the Kolmogorov-Smirnov method to test if a sample comes from a population with a specific distribution. The Kolmogorov-Smirnov test is distribution-free, in the sense that the critical values do not depend on the underlying distribution. The Anderson-Darling, proposed by Stephens (1974), is a modified Kolmogorov-Smirnov test that assigns more weight to the distribution's tails. The Cramer-von Mises test is similar to the Kolmogorov test, but it is more computationally complex (Thode 2002). Since the test results of these three methods on our data were very similar, for brevity, we only report the results of the Kolmogorov-Smirnov test (Table 2).

To address the concern that a small proportion of products are not normal, even though the majority of the products have normally distributed reviews, we conducted the normal test at the individual product level. Our results demonstrate that the distribution of online product reviews is not normal for almost 100% of the products. These results suggest that virtually all online product reviews in our sample do *not* have a normal distribution.

Traditionally, scholars have used parametric methods to test the homogeneity of a population or to determine the number of modes in a density function. For example, Everitt and Hand (1981) approximated the sampling distribution by a parametric mixture of unimodal distributions to assess goodness of fit (Aitkin and Rubin 1985, Roeder 1994). A disadvantage of this approach is that such assessment may be influenced by the validity of the particular parametric model and the hypothesis of homogeneity (Cheng and Hall 1998). This is because *testing* the goodness of fit along with *using* the goodness of fit can be very awkward (Cheng and Hall 1998).

Many non-parametric tests of modality have been proposed to address this issue. This includes the DIP test (Hartigan and Hartigan 1985), the excess mass test (Muller and Sawitzki 1991), and the Silverman (1981) test. The DIP and excess mass tests are equivalent in the one-dimensional case, in the sense that the excess mass statistic is exactly twice the DIP statistic (Cheng and Hall 1998). However, compared to the Silverman test, the DIP test is simpler, yet more conservative (Cheng and Hall 1998). The asymptotic level of the DIP test is zero for each non-zero value at nominal levels. Hence, it will be more likely to accept the unimodality using the DIP test. For example, Henderson et al. (2002) used many methods to test if worldwide per-capital incomes are converging.

---

[5] Bimodality is not due to a truncated distribution since consumers cannot write reviews higher than 5 or lower than 1 star. Graphic plots of the mean of product reviews other than 3.0 or 2.5 stars reveal that there are fewer consumers writing a review of 5-stars than those writing a 4-star review. In those cases, however, there are still fewer consumers leaving a review of 3-stars.

The authors found that during the years in which the DIP test did not yield multimodality, the other methods suggested more than two modes. Henderson et al. (2002) also noted that relative to the other competing tests, the DIP test requires a greater mass to signify a mode, while the additional modes may have gone unnoticed by the DIP test. In sum, there is substantial proof that the DIP test is a more conservative statistical test.

For these reasons, we selected the DIP test to formally test the assumption of unimodality on each individual product to examine whether the distribution of online product reviews is bimodal. The DIP test for a unimodal distribution approaches zero, while the DIP of a bimodal distribution approaches a positive constant.[6] If the DIP test shows a large percentage of product reviews to follow a bimodal distribution, our results are likely to be even more pronounced compared to the other, less conservative tests.

We obtained the DIP statistics using *R*, a language for statistical computing developed at Bell Laboratories (http://www.r-project.org). *R* offers a wide variety of statistical tests, such as nonlinear modeling,, time-series analysis, classification, and clustering. There are many examples of *R* being used for large-scale statistical data analysis, such as election forecasting in Austria, processing of gene data, and time-series analysis of brain images, among others (Ripley 2001).[7] For technical details on the DIP test and algorithm, please see Appendix 1.[8]

Similar to the normal test, we conducted the DIP test at the individual product level (Figure 4). As shown in Figure 4, for virtually all products with a mean review score between 1.5 and 4, their underlying distributions are *not* unimodal. Even for products with a mean product review score around 5 stars, 57.85%, 73.42% and 74.28% of the products in the Book, DVD and Video categories do not follow a unimodal distribution.

Figure 4: The Distribution of the Percentage of Rejection for Unimodal Distribution (95% Confidence Interval)

---
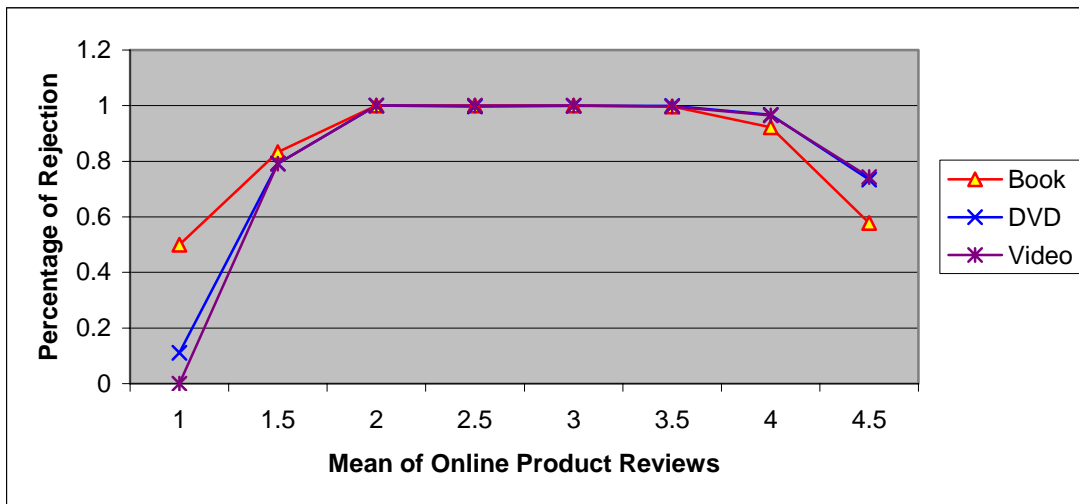
[6] The DIP test is a measure of departure from unimodality. This test measures the maximum difference between the empirical kernel distribution function from the unimodal distribution function that minimizes the maximum difference. In other words, the DIP is the distance between the "tightest fitting" unimodal distribution function and the empirical kernel distribution.
[7] For more information on the applications of *R*, please see Fox (2002) and Dalgaard (2002).
[8] For more detailed information on how to use *R* to compute the DIP values, please see Appendix 1, and also refer to: http://cran.r-project.org/doc/packages/diptest.pdf.

We summarize our normality and DIP test results in Table 2. As shown in Table 2, 90.17% of the products have a distribution that is neither normal nor unimodal. Only 9.83% of the products have a unimodal distribution. However, even for these products, their distribution is *not* normal.

Table 2: DIP Test and Normal Test Statistic at 95% Confidence Interval

DIP Test (Unimodal Test) / Normal Test	Accept	Reject	Total
Accept	0.00%	9.83%	9.83%
Reject	**0.00%**	**90.17%**	**90.17%**
Total	0.00%	100%	100%

To further establish bimodality, we tested the following quadratic equation (Anderson 1998):

$$f_{ijt} = \alpha_0 + \alpha_1 s_{ijt} + \alpha_2 s_{ijt}^2 + \sum \beta_m x_{jt,m} + \varepsilon_{ijt}, \tag{1}$$

where f_{ijt} is the number of product reviews with score i, for item j, at period t, $s_{ijt} \in \{1,2,3,4,5\}$ is review score i, for product j, at period t, ε_{ijt} is an error term. The null hypothesis to accept the bimodal distribution is given by:

$$H_0: \alpha_1 < 0 \text{ and } \alpha_2 > 0.$$

To account for potential differences in product attributes and different means, we ran a fixed effect model by regressing the number of product reviews on the number of stars by controlling for product category and the mean of online product reviews (Table 3). We also run a separate regression model for each product category with different mean of product reviews, and then estimated the mean coefficient across different groups.[9]

Table 3 shows the results when all products are pulled together. The results hold for books, DVDs, and videos with a significant negative $\alpha_1 < 0$ (-40.537), and also with a significant positive value $\alpha_2 > 0$ (9.088). These results also show that these online product reviews have a bimodal distribution. The results also hold for the regression model with all products and all sub-groups pooled together.

Table 3: Regression Results for Books, DVDs, and Videos

Items	Coefficient	p-value
Number of Stars	-40.537	0.0001
(Number of Stars)²	9.088	0.0001
Adj-R²	0.0429	
N	132978	

Discussion of Econometric Study

The econometric results prove that the online product reviews for the majority of products sold on Amazon do not follow a unimodal (or normal) distribution. In contrast, our econometric results suggest that the distribution of online product reviews on Amazon is bimodal, asymmetric, and positively skewed (J-shaped distribution). Even if there are no formal tests to prove that a distribution is bimodal, asymmetric, and positively skewed (J-shaped), the following findings give strong evidence for the existence of a J-shaped distribution of online product reviews:

---

[9] Since our condition index is smaller than 15, multicollinearity is not a concern. Also, given the positive correlation between the linear and quadratic terms, our hypotheses test when multicollinearity does not exist will be even stronger.

1. Online product reviews on Amazon are generally positive. As shown in Table 1, the mean of the reviews for the three focal product categories (books, DVDs and videos) is greater than 3.99 stars (on a 5-star scale). Also, as shown in Figures 2 and 3, there are significantly more positive (4- and 5-star) reviews than negative (1- and 2-star) reviews for all products in all three categories.

2. At the individual product level (excluding products with fewer than 20 reviews), 94.9% of books, 90.5% of DVDs and 90.3% of videos have a higher number of 5-star than 1-star reviews, further supporting the asymmetric, positively-skewed distribution of online product reviews.

3. Online product reviews tend to have more extreme than moderate reviews. As shown in Figure 2a, over half of the reviews are 5-star, 10% are 4-star, 10% are 3-star, 10% are 2-star, while 20% are 1-star reviews. This is also true for product with more than 2000 reviews in each category (Figure 2b), and also for products with a mean review score of 3 in each category, respectively (Figure 3). These findings support the J-shaped distribution of online product reviews.

Taken together, since the distribution of online product reviews is asymmetric, bimodal, and positively-skewed (J-shaped), the mean is a biased estimator. Therefore, the mean of a product's reviews does not reflect the aggregate evaluation of product quality; instead, it is a compromise among mainly polarized (1 or 5 stars) reviews. We offer three potential explanations for the observed bimodal J-shaped distribution of online product reviews:

- First, consumers do have heterogeneous (polarized or extreme) tastes, and the bimodal distribution simply describes the true state of nature that consumers indeed have polarized and not normally distributed reviews, while the J-shaped distribution indicates that consumers have more positive product evaluations. However, this also suggests that virtually all products in the market are outstanding, which is a weak argument (Kadet 2007).

- Second, consumers may be "overconfident" (Admati and Pfleiderer 2004) in their reviews, and the bimodal distribution is the result of consumer exaggerating (thus inflating) their polarized versus their moderate reviews.

- Third, consumers who purchase the product are already positively predisposed toward the product, and they are likely to report a positive review (purchasing bias). Also, consumers have heterogeneous incentives to report product reviews, and those with polarized reviews are more likely to assume the effort and time to write a review to "brag-or-to moan" compared to consumers with moderate reviews (under-reporting bias).

To help us identify which of these three competing explanations is the most likely, we performed a controlled lab experiment in which all consumers were asked to write their product reviews, as described below:

SECTION 2. EXPERIMENTAL STUDY

Description of Experimental Study

Since there is no direct way to obtain reviews from all consumers on Amazon, we performed a controlled lab experiment to obtain the distribution of product reviews when all consumers in the sample were asked to write reviews. We distributed a randomly-selected "Mr. A-Z" music CD to a sample of 66 students who were asked to listen to the CD and report their reviews on a custom-made website that closely resembled Amazon's website. 60 responses were obtained (91% response rate), implying that the product was reviewed by virtually all respondents.

Comparison between Experimental and Amazon's Online Product Reviews

The distribution of the experimental product reviews (Figure 5a) can be adequately fitted with a normal curve.

Figure 5a: Distribution of ratings in the experiment for a Music CD (Mr. A-Z)
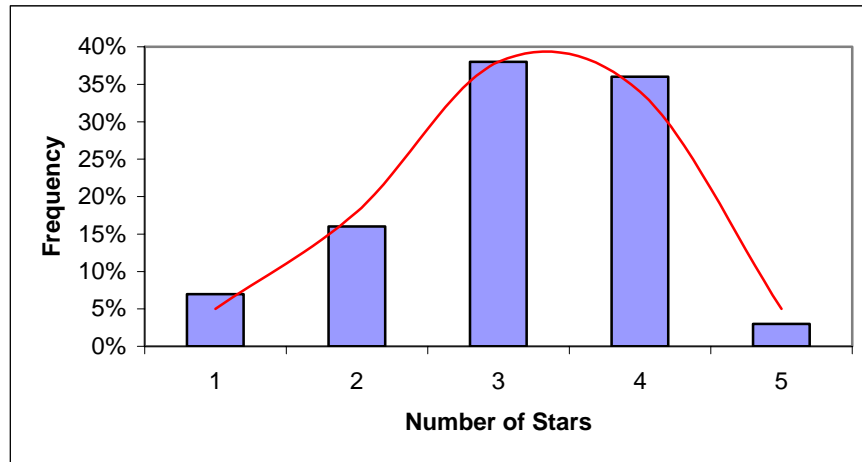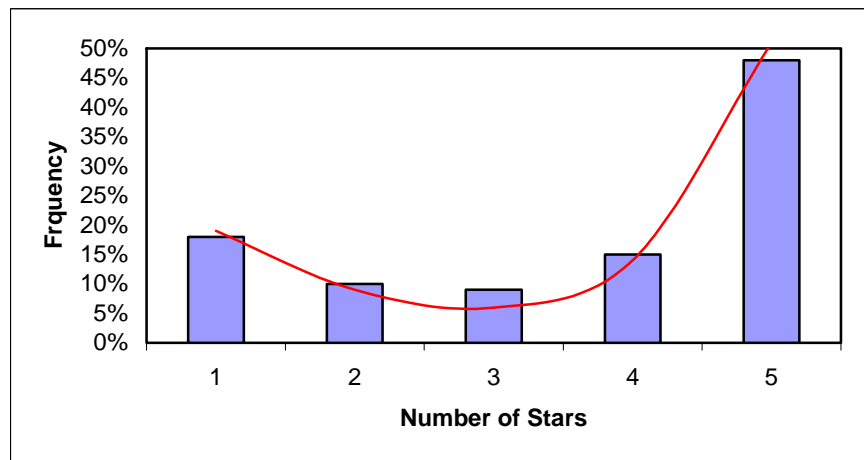**(Fitted with a Normal (Gaussian) Distribution)**

Figure 5b shows the distribution of Amazon's voluntary online product reviews for the "Mr A-Z" music CD. As shown in Figure 5b, the distribution fits a bimodal (J-shaped) distribution.

Figure 5b: Distribution of Amazon's Online Product Reviews for the "Mr A-Z" Music CD
**(Fitted with a Bimodal (J-shaped) Distribution)**

If Amazon's product reviews were representative of all consumers, Figures 5a and 5b would be comparable. However, the experimental results when all consumers reported their product reviews show a normal distribution (Figure 5a), whereas Amazon's results (Figure 5b) for the exact same product show a J-shaped distribution. Also, the mean of Amazon's online product reviews (mean=3.54) is significantly higher (p<.05) than the mean reported by the respondents of the experiment (mean=3.12). This different is attributed to the fact that Amazon's buyers voluntary purchased this product, and they are thus more likely to be positively predisposed to this product compared to a random sample of consumers. In fact, almost half of Amazon's reviews are 5-star reviews, while only 17% are 1-star reviews. This is because consumers purchase a product if they expect to derive utility from it (Chevalier and Mayzlin 2006), thereby creating a *purchasing bias* toward the product.

Also, the majority of Amazon's product reviews (>60%) are polarized (either 1-star or 5-stars), while less than 40% of the reviews is moderate (2, 3, or 4 stars). In contrast, the majority of the experiment's product reviews (over 90%) are moderate (2-star, 3-star, or 4-star), while less than 10% of the experimental reviews are polarized (1-star or 5-star reviews). These results show the sharp contrast between Amazon's voluntary and the experiment's involuntary reviews, suggesting that there is *under-reporting bias* when the reviews are reported voluntarily.

We also asked the respondents to rate whether they would report their product review on Amazon, and also how "passionate" they are with their reviews. Figure 6 shows the relationship between the respondents' "passion" about their product review and their willingness to voluntarily write a review on Amazon.

Figure 6. Relationship between Product Review Score and Willingness to Write Product Reviews on Amazon
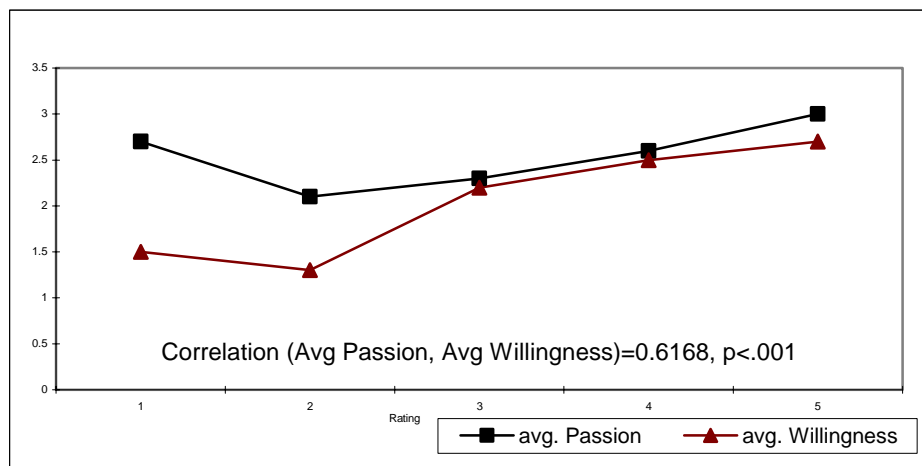
As shown in Figure 6, there is a significant correlation ($r = 0.62$, $p<0.001$) between the consumer's "passion" and their willingness to write a review on Amazon. This suggests that the more passionate consumers are, the more polarized their reviews are, and the higher is the likelihood to visit Amazon to voluntarily write a review. These results further strengthen the case of under-reporting bias since only passionate consumers intend to exert the effort to report their reviews, while less passionate consumers are less likely to report their moderate reviews.

Study Limitations

First, since the experimental results are based on a single product, there may be concerns about the generalizability of the experimental results to other products. While the chosen "Mr. Z" music CD is a standard music CD that was randomly selected, future research can attempt to reproduce the results to other products.

Second, the experiment was conducted among students, casting potential doubt to the generalizability of the results to other populations. Even though our students' age ranges from 18 to 40 years old that closely resembles the population of both the typical Amazon's consumers and buyers of music CDs, future research could replicate our experiment with a random sample of Amazon's consumers, buyers of music CDs, or consumers in general.

Third, while Amazon's product reviews are written by consumers who purchased the product, the respondents in our experiment did not actually purchase the product. While this difference also helps us infer the existence of purchasing bias, future research could replicate our experiment with consumers who voluntarily purchased a given product, and compare their "involuntary" reviews with Amazon's voluntary product reviews.

Discussion of Experimental Results

The experimental results help identify the most likely explanation for the J-shaped distribution observed in the voluntary Amazon reviews. One possible explanation is that *all* consumers do have polarized tastes, and their polarized reviews actually reflect the true state of nature. However, the experimental results show that the consumer reviews follow a roughly normal distribution, implying that most consumers do have moderate tastes. The second rival explanation was that consumers are overconfident in their reviews (Admati and Pfleiderer 2004), and they tend to exaggerate their polarized reviews. However, the experimental results suggest that only 10% of the consumers have polarized reviews for a randomly-selected product. Specifically, only 3% of the respondents in our experiment gave the music CD a 5-star review, and 7% of the respondents gave the product a 1-star review. Our third possible explanation is that consumers who purchased the product have higher product valuations, and they are likely to be more positively biased toward the product (purchasing bias), and also consumers with more polarized reviews are more likely to report their reviews to "brag-or-moan" versus those with moderate views (under-reporting bias). Integrating the econometric (Section 1) and experimental results (Section 2), we argue that the two independent sources of self-selection biases jointly give rise to the bimodal, J-shaped distribution.

SECTION 3. PRODUCT QUALITY AND FUTURE SALES PREDICTION MODELS

Our proposed "brag-or-moan" model aims to analytically derive the conditions for the mean to become an unbiased estimator of product quality, as described in detail below:

The "Brag-or-Moan" Analytical Model

Assume a product has an intrinsic quality q that follows uniform distribution on the support of 0 to 1, that is: $q : U[0,1]$. This quality variable q is a measure of the value of the product, which is unobservable before a consumer actually uses the product, but its distribution is common knowledge to all consumers. Besides this intrinsic quality, each consumer also experiences an idiosyncratic shock v_i that incurs when consuming a product; for example, taste, mood, social status, and other factors may affect consumer utility. Without loss of generality, we assume that this shock v_i is independent and identically normally distributed across consumers with mean 0 and variance σ^2, $v_i : N(0,\sigma^2)$. Then, consumer i's utility is given by:

$$u_i(q) = q + v_i. \tag{2}$$

Thus, given a product with quality q, the distribution of consumer i's utility is $u_i(q) : N(q,\sigma^2)$. If the price of the product is p, then only those consumers whose utility is greater than the price $u_i(q) \geq p$ will purchase the product. A consumer who purchases a product may or may not write an online product review. The benefit of writing a review is not obvious, but it does incur a cost (time and effort). Based on our previous results, we assume that the incentive for consumers to write a product review is to *brag* or *moan*. Following the marketing (Anderson 1998) and consumer behavior (Hoyer and MacInnis 2001) literature, there are psychological reasons to brag or moan. This is because consumers engage in WOM communication to express their satisfaction or dissatisfaction with a product, and to encourage or discourage others from purchasing the product. Anderson (1998) found that

consumers are more likely to engage in WOM communication if they feel strongly (either positively or negatively) about a product, rather than if they are indifferent. Therefore, we propose the function of the probability any given consumer chooses to write a review ρ to be influenced by both purchasing and under-reporting bias:

$$\rho = \begin{cases} 1 & \text{if } p < u_i(q) < \underline{\delta} \text{ or } \overline{\delta} < u_i(q); \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

where $\underline{\delta}$ and $\overline{\delta}$ are two cutoff values in a consumer's evaluation. Any value between them will not motivate consumers to write product reviews. Consumers with $u_i(q) < p$ will not purchase, and they will therefore not write a review. Assume both $\underline{\delta}$ and $\overline{\delta}$ are constant and $0 \le \underline{\delta} \le \overline{\delta} \le 1$. If $\underline{\delta} = \overline{\delta}$, it means that all consumers who made a purchase would write a review; otherwise, it implies that only extremely satisfied or extremely disgruntled consumers would write a review. Since we observe 1-star scores for almost all products in our data sample, we assume that $p \le \underline{\delta}$, which means that there will be at least some low-valuation consumers who purchase and write reviews. We also assume that every consumer who reports her product review score always truthfully reports her utility, as described in Equation (2), that is, $r_i = u_i(q)$ for $i \in \{i \mid p < u_i(q) < \underline{\delta} \text{ or } u_i(q) > \overline{\delta}\}$. We use $\overline{r}_n$ to represent the mean of the online product reviews reported by n consumers.

As consumers write online product reviews, they contribute information about product quality to the market. In the case that $\underline{\delta} = \overline{\delta}$ and price is very low where everyone purchases and reports her observed product quality, by the Central Limit Theorem we have:

$$\lim_{n \to \infty} \Pr[\frac{n^{1/2}(\overline{r}_n - q)}{\sigma} \le x] = \Phi(x), \tag{4}$$

that is, the distribution of the average product reviews $\overline{r}_n$ will be approximately a normal distribution with mean q and variance σ^2/n. Thus, consumers can indeed learn a product's true quality from the online product reviews.

More generally, when $p < \underline{\delta} < \overline{\delta}$, the distribution of reported quality r_i is a normal distribution $N(q, \sigma^2)$ but with a window as described in Equation (3). Our analysis leads to the following proposition regarding whether the mean of the online product reviews can be used as an unbiased estimator of product quality:

Proposition 1: When the number of product reviews n is large enough, the mean of online product reviews $\overline{r}_n$ converges to $q + \Lambda\sigma$, and the variance σ_r^2 converges to $(1 + \Lambda - \Lambda^2)\sigma^2 - \Lambda q\sigma$, where

$$\Lambda = \frac{\phi(\frac{\overline{\delta}-q}{\sigma}) - \phi(\frac{\underline{\delta}-q}{\sigma}) + \phi(\frac{p-q}{\sigma})}{(\Phi(\frac{\underline{\delta}-q}{\sigma}) - \Phi(\frac{p-q}{\sigma})) + (1 - \Phi(\frac{\overline{\delta}-q}{\sigma}))}.$$ **The condition for the mean of online product reviews $\overline{r}_n$ to be an**

unbiased estimator of the mean product quality q is: $\phi(\frac{\overline{\delta}-q}{\sigma}) = \phi(\frac{\underline{\delta}-q}{\sigma}) - \phi(\frac{p-q}{\sigma})$.

Proof: see Appendix 2.

The purchasing bias inflates the mean of the product reviews since low valuation consumers are screened out, and have no chance to write product reviews. Even if we take out the purchasing bias and assume that everyone purchases, then the condition in Proposition 1 is reduced to: ($\underline{\delta} = \overline{\delta}$) or ($\underline{\delta} < \overline{\delta}$ and $0 \le q - \underline{\delta} = \overline{\delta} - q$), which would also require either all consumers to write a review (similar to the experiment in Section 2), or to have a symmetry between the extremely satisfied and extremely disgruntled consumers who report their product reviews. Since both conditions are unlikely to be satisfied, the mean of the online product reviews based on a truncated sample is still unlikely to represent product quality. More specifically, if $\phi(\frac{\overline{\delta} - q}{\sigma}) < \phi(\frac{\underline{\delta} - q}{\sigma}) - \phi(\frac{p - q}{\sigma})$, that is, consumers are more likely to moan (satisfied consumers are less likely to write reviews), the mean of the reviews will be lower than "true" product quality; otherwise, if consumers are more likely to brag (dissatisfied consumers are less likely to write reviews), the mean of the online product reviews will be higher than "true" product quality.

The proposed model suggests that a consumer's biases to write an online product review have important implications on whether online product reviews can be used as a good estimator of product quality. We show that the commonly-used factor in the literature – the mean of the online product reviews - is likely to be a biased estimator for product quality. Therefore, making this unwarranted assumption can possibly lead to erroneous conclusions about consumer assessments of product quality, leading to poor marketing decisions for managers.

Proposition 1 also suggests that we can find an unbiased estimator of product quality q and standard deviation of consumer types σ by solving the simultaneous equation:

$$\begin{cases} \overline{r}_n = q - \Lambda\sigma \\ \sigma_r^2 = (1 + \Lambda - \Lambda^2)\sigma^2 - \Lambda q\sigma \end{cases}. \tag{5}$$

Therefore, the first equation in (5) indicates that quality q is a function of the mean of the product reviews $\overline{r}_n$, standard deviation σ, and Λ. A consumer observes the mean of the product reviews $\overline{r}_n$ and subtracts $\Lambda\sigma$ in order to overcome the two types of biases, and properly estimate the product quality. Without under-reporting bias:

$q = \overline{r}_n - \dfrac{\phi(\frac{p-q}{\sigma})}{1 - \Phi(\frac{p-q}{\sigma})}\sigma$, where $\dfrac{\phi(\frac{p-q}{\sigma})}{1 - \Phi(\frac{p-q}{\sigma})}\sigma$ is used to rectify the purchasing bias. Without purchasing bias:

$q = \overline{r}_n - \dfrac{\phi(\frac{\overline{\delta}-q}{\sigma}) - \phi(\frac{\underline{\delta}-q}{\sigma})}{\Phi(\frac{\underline{\delta}-q}{\sigma}) + (1 - \Phi(\frac{\overline{\delta}-q}{\sigma}))}\sigma$, where $\dfrac{\phi(\frac{\overline{\delta}-q}{\sigma}) - \phi(\frac{\underline{\delta}-q}{\sigma})}{\Phi(\frac{\underline{\delta}-q}{\sigma}) + (1 - \Phi(\frac{\overline{\delta}-q}{\sigma}))}\sigma$ is used to address the under-reporting bias.

The simultaneous Equation (5) contains Λ, which is defined in a complicated form of product quality q, the standard deviation of all consumers' utilities σ, and the utility bounds $\underline{\delta}$ and $\overline{\delta}$ within which consumers do not prefer to report their product reviews. Therefore, the solution to Equation (5) has no closed-form solution for product quality and the standard deviation of all consumers' utilities, and it can only be estimate numerically. Our

15

proposed analytical model suggests that an unbiased estimator of product quality is a function of: (1) the mean of the product reviews $\bar{r}_n$, (2) the standard deviation σ_r, (3) the two utility bounds $\underline{\delta}$ and $\overline{\delta}$ (which are used to account for under-reporting bias), and (4) also product price p (which is used to account for purchasing bias). Therefore, we propose that a model that incorporates the mean of the online product reviews $\bar{r}_n$, the standard deviation σ_r and the two utility bounds $\underline{\delta}$ and $\overline{\delta}$, and product price p would be a superior model than the one with only the mean of the online product reviews.

Given a product with quality q and price p, only consumers with utility greater than or equal to the price p will purchase the product. Therefore, the product demand generated by consumers is given by:

$$D(q) = \int_{p-q}^{\infty} d\Phi(v_i) = 1 - \Phi(p - q) \,. \tag{6}$$

Since the quality q is not directly observable from Amazon's website before purchasing, consumers are inferring product quality based on the online product reviews according to Equation (5). Plugging $q = \bar{r}_n - \Lambda\sigma$ to the demand function (Equation 6), we propose the following comparative static results:

Proposition 2:

(i) **The demand of a product is positively correlated with the mean of its online product reviews $\bar{r}_n$, and negatively correlated with its price p.**

(ii) **When there exists a significant under-reporting bias, (both $\overline{\delta} - q$ and $q - \underline{\delta}$ are not too small), product demand is positively correlated with the upper-bound of utility for those consumers who do not write online product reviews $\overline{\delta}$, and negatively correlated with the lower-bound of utility for those consumers who do not write online product reviews $\underline{\delta}$.**

(iii) **The standard deviation of observed ratings σ_r is correlated with product demand, and the degree of correlation depends on the level of the purchasing bias and the under-reporting bias.**

Proof: See Appendix 2.

Proposition 2 suggests that, ceteris paribus, a higher mean of online product reviews will indicate a higher product quality and therefore increase product demand. Product demand is also influenced by the two modes and the standard deviation of the distribution of online product reviews. A price increase has a negative effect on product demand by reducing the number of consumers with higher product valuations who can afford the product.

The Proposed 'Distribution-Based' Product Sales Forecasting Model

Ceteris paribus, a higher product quality increases consumer utility and product demand, resulting in higher future product sales. Hence, 'future product sales' is proposed as a meaningful and tangible proxy for reflecting product quality. Sales forecasting models based on product reviews has been studied by Clemons et al. (2004), Duan et al. (2005), Liu (2006), Sorensen and Rasmussen (2004), among others. Based on our analytical model, we propose a 'distribution-based' model to predict future sales by integrating the following variables: (1) the mean of the online product reviews $\bar{r}_n$, (2) the standard deviation σ_r, (3) the two utility bounds $\underline{\delta}$ and $\overline{\delta}$, (4) and price p.

Based on Proposition 2, we propose the following four hypotheses:

H1: The mean of online product reviews $\bar{r}_n$ is *positively* associated with future product sales.

H2: The standard deviation of the online product reviews σ_r is *negatively* associated with future product sales.

H2: The lower mode $\underline{\delta}$ of online product reviews is *negatively* associated with future product sales, while the upper mode $\bar{\delta}$ is *positively* associated with future product sales.

H4: Product price *p* is *negatively* associated with future product sales.

While Proposition 2 specifies the directionality of the relationship between the mean, each of the two modes, and product price, it does not specify the directionality of the relationship between the standard deviation of the online product reviews with product demand. A high standard deviation suggests that there many polarized product reviews, and that the two modes of the distribution are far apart. This makes it difficult for consumers to infer product quality, creating a high degree of uncertainty for consumers, which is likely to negatively influence their product demand and future product sales. This argument is consistent with Clemons et al. (2004) who show that the variance of product reviews negatively affects future product sales. Therefore, we hypothesize that the standard deviation of online product reviews to be negatively associated with future product sales (H2).

We compare the proposed distribution-based model with competing models based on how well they predict future sales in terms of variance explained (R^2 Adjusted). We propose three models with single-point estimators: (1) the simple mean of the online product reviews (Model 1); (2) three distinct *weighted* mean scores (Model 2), and (3) a model that uses the percentage of polarized (1-star or 5-star) product reviews (Model 3), following Chevalier and Mayzlin (2006). First, the simple mean is the easiest proxy for inferring product quality, and it is thus used as a potential predictor of future sales. Second, the weighted mean of the online product reviews is used because consumers may weigh product reviews differently. Based on the relative relationship among product reviews, helpful reviews, and total reviews, we propose three different weighted mean measures: (a) mean weighted by the percentage of consumers who think the review is helpful; (b) mean weighted by the sum of percentages of consumers who think it is helpful across reviews; and (c) mean weighted by the number of consumers who think the review is helpful. Finally, we also include the percentage of polarized (1-star or 5-stars) reviews since consumers may pay greater attention to them (e.g., Anderson 1998, Chevalier and Mayzlin 2006). Please note there are two density masses in the distribution of product reviews for polarized positive and negative reviews $-X_L$ and X_U - which are different from the percentage of polarized (1-star and 5-star) product reviews.

We formally estimate the two modes $\underline{\delta}$ and $\bar{\delta}$ with the lower and upper ends (X_L and X_U) of the modal interval returned by the DIP test (Hartigan and Hartigan 1985). After inputting the observed online product reviews, the statistical package **R** can generate the required DIP statistics (Hartigan 1985) $-X_L$ and $X_U -$ which can serve as proxies for $\underline{\delta}$ and $\bar{\delta}$, respectively. X_L and X_U represent the centers of the two modes of the online product reviews. As noted in Section 1 and described in Appendix 1, the DIP test is a measure of departure from unimodality by

measuring the maximum distance between the kernel distribution function from the "tightest fitting" unimodal distribution function that minimizes this maximum distance. The DIP test thus provides the centers of the modes.

Since product sales rank is shown in a descending order where 1 represents the best selling product, there is a negative correlation between product sales and sales rank (sales rank is directly available by Amazon). As shown by many authors (e.g., Brynjolfsson, Hu and Smith 2003, Chevalier and Goolsbee 2003, Ghose *et al*. 2006), the relationship between the sales rank and the actual volume of product sales on Amazon can be approximated by:

$$\ln(Sales) = \beta_0 - \beta_1 * \ln(SalesRank).$$

Similar log-linear relationships have been documented in the literature (e.g., Schnapp and Allwine 2001). Pareto (1896) proved that income could be approximated by a log-linear distribution. Zipf (1949) showed that city size followed a log-linear distribution with a slope of -1. Given the linear relationship between ln(*Sales*) and ln(*SalesRank*), we use ln(*SalesRank*) as a proxy for future product sales in our log-linear regression models.

The following five models were compared on their predictive validity:

Model X (Proposed 'Distribution-Based' Model)

$$\ln(SalesRank_{i+1}) = \beta_{01}\ln(Mean\_Rating_i) + \beta_{21}X_{Li} + \beta_{31}X_{Ui} + \beta_{41}\ln(Stdev\_Rating_i)$$
$$+ \beta_{51}\ln(SalesRank_i) + \beta_{61}\ln(Price_i) + \beta_{71}\ln(Num\_Rev_i)$$
$$+ \beta_{71}Book\_Dummy + \beta_{81}DVD\_Dummy + \varepsilon_{ix}$$

Model 1 (Simple Mean Model)

$$\ln(SalesRank_{i+1}) = \alpha_{01} + \alpha_{11}Mean\_Rating_i + \alpha_{21}\ln(SalesRank_i) + \alpha_{31}\ln(Price_i) + \alpha_{41}\ln(Num\_Rev_i)$$
$$+ \alpha_{51}Book\_Dummy + \alpha_{61}DVD\_Dummy + \varepsilon_{i1}$$

Model 2a (Weighted Mean a*)

$$\ln(SalesRank_{i+1}) = \alpha_{02} + \alpha_{12}Weighted\_Mean\_Rating_{1i} + \alpha_{22}\ln(SalesRank_i) + \alpha_{32}\ln(Price_i) + \alpha_{42}\ln(Num\_Rev_i)$$
$$+ \alpha_{52}Book\_Dummy + \alpha_{62}DVD\_Dummy + \varepsilon_{i2}$$

Model 2b (Weighted Mean b*)

$$\ln(SalesRank_{i+1}) = \alpha_{03} + \alpha_{13}Weighted\_Mean\_Rating_{2i} + \alpha_{23}\ln(SalesRank_i) + \alpha_{33}\ln(Price_i)$$
$$+ \alpha_{43}\ln(Num\_Rev_i) + \alpha_{53}Book\_Dummy + \alpha_{63}DVD\_Dummy + \varepsilon_{i3}$$

Model 2c (Weighted Mean c*)

$$\ln(SalesRank_{i+1}) = \alpha_{04} + \alpha_{14}Weighted\_Mean\_Rating_{3i} + \alpha_{24}\ln(SalesRank_i) + \alpha_{34}\ln(Price_i) + \alpha_{44}\ln(Num\_Rev_i)$$
$$+ \alpha_{54}Book\_Dummy + \alpha_{64}DVD\_Dummy + \varepsilon_{i4}$$

Weighted_Mean_Rating$_{1i}$ = Avg((Helpful#/Total#) Rating)
Weighted_Mean_Rating$_{2i}$ = Sum((Helpful#/Total#) Rating)/Sum(Helpful#/Total#)
*Weighted_Mean_Rating$_{3i}$ = Sum(Helpful#*Rating) / Sum(Helpful#)

Model 3 (1-Star & 5-Star Model)

$$\ln(SalesRank_{i+1}) = \alpha_{05} + \alpha_{15}Percent_{1star} + \alpha_{25}Percent_{5star} + \alpha_{35}\ln(SalesRank_i) + \alpha_{45}\ln(Price_i)$$
$$+ \alpha_{55}\ln(Num\_Rev_i) + \alpha_{65}Book\_Dummy + \alpha_{75}DVD\_Dummy + \varepsilon_{i5}$$

To compare the proposed 'distribution-based' model against these five competing models, we randomly selected another 10,000 books, DVDs, and videos from our Amazon sample in July 2005 using the SAS random function. For each product, we collected its price, sales rank, and online product reviews from July 2005 to January 2006. These longitudinal secondary data are used to compare which model has the highest power in terms of predicting future product sales using panel data. We define the duration of each period as 130 days. We also tested other time lag values (e.g., 100 days, 110 days), which yielded similar results.

As shown in Table 4, all four hypotheses are supported. Controlling for previous sales rank, price, and the total number of online product reviews, the proposed 'distribution-based' model explains a substantial amount of the variance in future product sales (R^2 adjusted=77.29%), which is significantly higher than all other models (p<.0001). The *mean* of the product reviews has a significant effect, explaining .16% of the variance in future product sales (supporting H1). The *standard deviation* of the product reviews is also statistically significant (p<.05), explaining .034% of the variance, supporting H2. This finding is consistent with Clemons et al. (2004) who show that the variance of product reviews affects future product sales. The two modes X_L and X_U *also* have significant effects (p<.001),[10] explaining .64% and .33% of the variance in future product sales, respectively, thus supporting H3. Interestingly, the variance explained by X_L (lower mode) is almost twice as much as that explained by the upper mode X_U. These findings are consistent with Chevalier and Mayzlin 2006) who showed that the impact of 1-star reviews was greater than the impact of the 5-star reviews on book sales on Amazon.com and Barnes & Noble.com. Finally, *product price* has a significant effect on future product sales, supporting H4.

Table 4: Model Comparisons (Dependent Variable: ln(Future Sales Rank))

	Proposed Model X	Model 1	Model 2a	Model 2b	Model 2c	Model 3
Mean_Product Reviews	-0.0513***	-0.0351***	-0.0232**	-0.0272**	-0.0194*	
Stdev_Reviews	0.0086*					
X_U	-0.0685***					
X_L	0.0573***					
ln(Price)	0.0780***	0.0999***	0.100***	0.0989**	0.0986***	0.099***
Percent(5-star rating)						-0.0452
Percent(1-star rating)						0.1088*
Ln (Current Sales Rank)	0.7114***	0.7206***	0.7218***	0.721***	0.7215***	0.7208***
ln (# of Rating)	-0.0869***	-0.0533***	-0.056***	-0.050***	-0.051***	-0.0528***
Book Dummy	0.311***	0.334***	0.333***	0.331***	0.390***	0.333***
DVD Dummy	-0.154***	-.1042***	-0.105***	-0.108***	-0.108***	-0.105***
Intercept	3.7890***	3.5234***	3.448***	3.482***	3.444***	3.3936***
Adjusted R²	**77.29%**	**75.18%**	**75.17%**	**75.25%**	**75.24%**	**75.17%**
Difference in Adjusted R²		**2.11%**	**2.12%**	**2.04%**	**2.05%**	**2.12%**
F-Value		23,444***	23,555***	22,666***	22,778***	23,555***
N	7573	7573	7573	7573	7573	7573

*** p<.001, ** p<.01, * p<.05, + p<.10. (All p-values are two-sided)

---

[10] Following Aigner (1971), the variance explained was decomposed among the independent variables by multiplying the standardized regression coefficients by the correlation of the independent variables with the dependent variable.

There are several criteria that can be used to choose among competing models, such as the R^2 Adjusted, Akaike information criterion (AIC), Schwarz Information criterion (SIC), Mallow's C_p criterion, and forecast χ^2 (chi-square). These criteria aim at minimizing the residual sum of squares, or increasing the adjusted R^2 value. The AIC imposes a harsher penalty than the R^2, while the SIC imposes an even harsher penalty than the AIC. However, as argued by Diebold (2001), no criterion is necessarily superior. For simplicity, we evaluated the performance of the various models by comparing their adjusted R^2, which is the most widely used criterion. Accordingly, we used the following equation for calculating the significance between two regression models:

$$F_{(kx-ki),(n-kx-ki)} = \frac{[R^2(Model\_X) - R^2(Model\_i)]/K_x - K_i}{[(1 - R^2(Model\_X)]/(N - K_x - K_i)} \qquad (7)$$

where: K_x is the number of independent variables in the proposed Model X

K_i is the number of independent variable in the competing Model i ($i = 1$, 2a, 2b, 2c, and 3)

N is the sample size.

In terms of other comparisons beyond the F-test (Equation 7), following Davidson and MacKinnon (1993, p. 456): "For linear regression models, with or without normal errors, there is of course no need to look at likelihood, W, and LR at all, since no information is gained from doing so over and above what is already contained in F." Therefore, we did not perform other comparisons for the nested models. Also, for non-nested model comparison, we used the Davidson-Mackinnon J test, which rendered similar results.

The proposed distribution-based model explains at least 2% higher variance compared to the five competing models (which roughly explain about the same variance).[11] This difference in variance explained is statistically significant (p<.001), as the F-tests in Table 4 attest. Besides the high F-values that show that the 2% improvement in variance explained is statistically significant, from a practical standpoint, one may question this improvement. However, it is important to recognize that the great majority of the variance is explained by the control variables. Specifically, the current sales rank explains 52% of the variance,[12] the number of product reviews explains 4.55%, the product dummy variables explain 7%, and price only explains .11%. Given these influential control variables, the variance explained by the new independent variables (X_L, X_U, $Stdev$) is substantial from a practical standpoint as well, attesting to the need for including these distributional parameters when predicting future product sales.

While the proposed prediction model with the proposed X_L and X_U modes is superior to the one with the polarized (1-star and 5-star) reviews (Model 3), we still wanted to have a direct comparison of their joint impact. Therefore, we run a regression model in which we included both the X_L and X_U and also the percentage of 1-star and 5-star reviews (Table 5). The density mass in the bimodal distribution of online product reviews that obtains the X_L and X_U are different from the percentage of 1-star and 5-star reviews, allowing us to simultaneously include them in a regression model. As shown in Table 5, both the percentage of polarized (1-star and 5-stars) reviews

---

[11] Interestingly, none of the three proposed weighted means of the online product reviews is superior to the simple mean. Perhaps this is because consumers only observe the simple mean and do not process the number of useful reviews.
[12] We also run the same regression models (Table 4) by omitting the current sales rank as a control variable, and the results were very similar (Model X outperformed all others by over 2% in adjusted R^2). However, since the variance explained in future sales is much lower (circa 35%) when omitting current sales rank, we only report the results with all control variables.

become insignificant when X_L and X_U are included in the regression model. Accordingly, the inclusion of polarized reviews did not improve the variance explained (77.29%) in future product sales. These results attest to the superiority of the X_L and X_U parameters obtained by the DIP test versus the percentage of polarized reviews.

Table 5: Comparison between Proposed Modes and Percentage of Polarized Reviews

	Regression Model
Mean_Product Reviews	-0.0906*
X_L	0.0563***
X_U	-0.0688***
Stdev_Reviews	0.0083*
Percent(5-star reviews)	-0.1895[N/S] (p=0.1739)
Percent(1-star reviews)	0.0343 [N/S] (p=0.684)
ln (Current Sales Rank)	0.71152***
ln(Price)	0.0778***
ln (# of Reviews)	-0.0866***
Book Dummy	0.318***
DVD Dummy	-0.152***
Intercept	3.95***
Adjusted R^2	**77.29%**
N	7573

DISCUSSION

Key Findings

First, econometric results from data drawn from the most popular online retailer Amazon.com show that the majority of online product reviews follow an asymmetric bimodal (J-shaped) distribution. These findings are supported with data about products from different categories and multiple means of online product reviews. While the literature has only speculated about the bimodal distribution of product reviews (Admati and Pfleiderer 2004), to the best of our knowledge, this is the first study to empirically show the asymmetric, bimodal, positively-skewed (J-shaped) distribution of online product reviews with longitudinal real-life data from the largest online retailer.

Second, experimental results help explain whether the bimodality of online product reviews is explained by consumer heterogeneity, overconfidence, or purchasing and under-reporting bias. By showing that the distribution of product reviews follows a normal distribution when all consumers are asked to write a review, the experimental results show that purchasing bias (due to self-selection of consumers who purchase the product and write reviews) and under-reporting bias (due to the suppression of moderate product reviews because such consumers have lower incentives to voluntary report their reviews) jointly explain the J-shaped distribution of online product reviews.

Third, we provide an explanation to the econometric and experimental findings by proposing an analytical model to derive the conditions under which the mean of the online product reviews can be an unbiased estimator of product quality. Our analytical model shows that the two self-selection (purchasing and under-reporting) biases cause an asymmetry in the number of consumers who write product reviews to "brag or to moan," rendering the mean of the online product reviews a biased estimator of product quality. The analytical model also derives the

three distributional parameters – (1) the mean, (2) standard deviation σ_r, and (3) the two utility bounds $\underline{\delta}$ and $\overline{\delta}$ (that jointly account for the under-reporting bias), and (4) product price p (that accounts for the purchasing bias) – that are needed in order to create an unbiased estimator of product quality.

Fourth, based on the analytically derived distributional parameters by which the mean can become an unbiased estimator of product quality, we propose a new model that is used to predict future product sales. The proposed 'distribution-based' model has a higher predictive power in terms of explaining the variance in future product sales compared to the simple and weighted mean and also the percentage of polarized (1-star and 5-stars) reviews.

Taken together, this study integrates econometric data with experimental findings and analytical modeling to question the underlying assumption in the literature that the mean of online product reviews is an unbiased estimator of product quality due to two self-selection biases (purchasing and under-reporting) that can be partially overcome by using a set of distributional parameters together with product price.

Implications for Theory and Research

The paper has implications for theory and research on: (1) the nature of online product reviews, (2) potential biases in online WOM communication, and (3) online consumer behavior in general, as we discuss below:

Implications for the Nature of Online Product Reviews

While there is an emerging interest in the nature and role of online product reviews (e.g., Chen *et al*. 2004, Chevalier and Goolsbee 2003, Chevalier and Mayzlin 2005, Duan *et al*. 2005, Godes and Mayzlin 2004), there is still no consensus as to whether, how, and why online product reviews can adequately predict product sales. Virtually all studies use the mean as a predictor of product sales product quality (besides two recent exceptions) (Chevalier and Mayzlin 2006, Liu 2006), assuming that it is an unbiased estimator of product quality. We challenged this assumption by bringing the nature of online product reviews into close scrutiny, and empirically, experimentally, and analytically showing that there are potential self-selection biases in online product reviews. The demonstrated J-shaped distribution of online product reviews is explained by purchasing bias (since only consumers with positive dispositions and high product valuations would purchase products and write reviews) and under-reporting bias (since consumers with moderate reviews do not have the same motivation to report their product reviews as consumers with polarized reviews). In fact, it is estimated that 1 out of 1000 consumers writes a product review on Amazon.[13] This explanation is empirically verified compared to alternative explanations of consumer heterogeneity and consumer overconfidence, confirming that purchasing and under-reporting biases are what mainly renders the mean as a biased estimator of product quality. Simply put, reporting bias causes a bimodal, U-shaped distribution, and purchasing bias causes the observed asymmetric, J-shaped distribution.

Our analytical model identifies the conditions for the mean of online product reviews to become an unbiased estimator of product quality. Since consumers have different incentives to voluntarily write product reviews, for the mean to become an unbiased estimator of product quality, we first need to account for product price to help overcome purchasing bias. We also need to include the mean, the standard deviation, and the two modes of the

---

[13] http://www.freakonomics.com/blog/2005/07/22/why-do-people-post-reviews-on-amazon/

distribution of online product reviews. Based on the conditions derived by our "brag-or-moan" analytical model, our proposed distribution-based model is shown to better predict future product sales compared to competing models that do not account for the bias in the mean. Since future product sales is a good proxy of product quality (defined as the collective assessment of a product's perceived value across consumers), our proposed model that introduces additional parameters from the distribution of online product reviews plus product price is shown to outperform several competing models when predicting future product sales.

Implications for Overcoming Biases in Online WOM Communication

Online WOM communication is becoming a popular informational source for consumers and marketers, and researchers are focusing their attentions on the impact of online WOM communication on consumer behavior and product success. However, the literature is impeded by the idiosyncrasies of online WOM communication that may give rise to self-selection biases that need to be accounted for:

First, research on online WOM communication should take into account the potential for under-reporting bias when single-point estimators are used, such as the simple or weighted mean of online product reviews and the percentage of polarized reviews to represent product quality. As this study attests, distributional parameters, such as the lower and upper ends of the modal interval returned by the DIP test can better reflect product quality. Thus, future research on online WOM communication must take in account the distribution of product reviews, and go beyond single-point estimators by incorporating additional distributional parameters. We argue that ignoring the distribution of online product reviews can potentially lead to invalid statistical assumptions and even erroneous conclusions about consumer product preferences, potentially resulting in incorrect marketing decisions.

Second, research on online WOM communication should also consider the potential for purchasing bias since product reviews are only written by consumers who voluntarily purchased the product. This is because only consumers with product valuation greater than or equal to price will actually purchase the product, and thus have the opportunity to write a product review. To account for purchasing bias, we analytically integrated product price into our proposed model. Therefore, future research on online WOM communication must also account for purchasing bias and also consider other variables, besides product price, to further overcome purchasing bias.

Implications for Online Consumer Behavior

The study's results have implications for online consumer behavior since the *observed* mean of consumer preferences is based on a truncated sample of the entire consumer population, resulting in a biased estimation of what consumers collectively perceive as product quality. However, our results imply that consumers are rational, and they take intuitive steps to overcome the self-selection biases in the nature of online product reviews. Since our model predicts *actual* future product sales, our results suggest that consumers incorporate the three conditions (mean, standard deviation, and modes of the polarized reviews) to form their product purchasing decisions.

First, consistent with the literature, consumers do rely on the mean of product reviews. This is intuitive since the mean is a figure that is readily observed and easily processed and understood.

Second, consumers also take in consideration how far spread apart the two modes of the distribution of online product reviews are (standard deviation of the bimodal distribution). The farther the two modes are (which can be

inferred as a proxy for the distribution's standard deviation), the higher the uncertainty in inferring product quality is, and the greater is the negative effect on product demand and future sales.

Third, consumers also take in account where the two modes are, and they weigh the negative (lower) mode more (roughly twice) than the positive (upper) model. This finding suggests consumers assign greater weight to negative WOM communication than to positive one, consistent with the consumer behavior literature that suggests that negative reviews are more influential than positive ones (e.g., Arndt 1967). The marketing literature has asserted that negative WOM communication is twice or even three times more influential than positive one (e.g., Herr Kardes, and Kim 1991, Mizerski 1982, Richins 1983) on the basis of being surprising or unexpected. However, we argue that negative WOM communication is more influential because negative reviews are under-represented compared to positive ones due to purchasing bias. Rational consumers recognize the purchasing bias, and they compensate for this bias by taking negative reviews more seriously and discounting the positive reviews. Our results suggest that negative consumer reviews are weighed roughly twice more heavily than positive ones.

These findings are very interesting since they suggest that consumers are "smart" enough to compensate for the under-reporting and purchasing bias by using additional information from the distribution of online product reviews to form the quality perceptions, shape their purchasing decisions, and create aggregate product demand. These findings are consistent with Banerjee and Fudenberg (2004) who argue that under-reporting bias does not inhibit social learning since smart consumers recognize that self-selection bias exists in online product reviews, and they use heuristics to compensate for the observed bias when forming their estimations of product quality.

Implications for Practice

This study shows that the mean of the online product reviews is a biased estimator of product quality due to self-selection biases. Assuming that the mean is an unbiased estimator of product quality can possibly lead to erroneous conclusions about consumer preferences and future product demand, leading to poor marketing decisions for managers. Hence, this paper has practical implications for online retailers, product manufacturers, consumers, and firms that specialize on the collection and dissemination of online product reviews.

First, while online retailers (e.g., Amazon.com, Buy.com and Circuitcity.com) encourage their consumers to write and read product reviews, the simple mean of the reviews they usually highlight seems to be a biased indication of a product's quality. However, the proposed distribution-based model can be used by online retailers to provide superior information to their consumers that will not require them to employ intuitive heuristics to overcome the purchasing and under-reporting biases in online product reviews. Specifically, using the DIP test, online retailers can identify where exactly the modes of the polarized reviews are. By having more accurate information on a product's future sales based on the distribution of a product's online reviews, online retailers can accordingly adjust their product supplies to take advantage of increased or reduced future product demand.

Second, the proposed model can also be used by retailers to predict a product's long-term success based on the reviews posted by early adopters. More specifically, the proposed distribution-based estimation model allows managers to more accurately predict which products are more likely to be successful in the future. In doing so, managers can more reliably estimate which products to stock, and thereby forge long-term arrangements with

their product suppliers. Accordingly, having a superior estimation of future consumer demand can also be used by product manufacturers and other players in a product's value chain to adjust their production and distribution schedules to focus on products that are likely to have higher or lower demand in the future.

Third, since many companies (e.g., Amazon, Barnes and Noble, Epinions.com and BizRate.com) specialize on the collection, synthesis, and dissemination of online WOM communication in the form of product reviews, having a more accurate method to inform consumers on aggregate product quality can be a differentiating factor. For example, tuning the reported mean by accounting for the two observed modes and the standard deviation can provide an unbiased mean value that can be directly used by all consumers to infer product quality and make better purchase decisions. Even if Banerjee and Fudenberg (2004) argue that "smart" consumers are able to mentally see through the "under-reporting bias" and infer the quality of products from biased product reviews, the proposed distribution-based estimation technique can alleviate this additional cognitive burden from consumers.

Limitations and Suggestions for Future Research

This study has several limitations that create interesting opportunities for future research:

First, even if the experimental study showed that the online product reviews of the focal product followed a roughly normal distribution, it is important to note that the experiment was conducted with a small sample size and a single product. Future research should replicate the experiment with a larger sample size and more products before concluding that product reviews have a normal distribution when all consumers write their product review.

Second, product price is not expected to solely overcome purchasing bias since many other factors are likely to deter consumers from purchasing a product. Future research could identify such factors that could help further overcome purchasing bias. Also, in order to make our analytical model more tractable, we assumed that moderate consumers in the middle of the distribution of online product reviews would not report a product review at all. Under this assumption, we derive an intuitive and meaningful condition for the mean of product reviews. However, future research could pursue a more natural approach that would assume a beta distribution for the probability of a moderate consumer to write a product review, and thereby derive a more precise condition.

Third, despite the study's longitudinal data, our econometric study did not consider the time-series properties of online product reviews, which can be assessed by future research. Also, since our study only uses data from Amazon, it might be useful to see whether our results will be similar in other retail websites, such as Epinions, eBay, Half.com, and other online retailers. Moreover, our results only use the numerical score (number of stars) of the consumer's online product review. However, consumers often write a more detailed product assessment with text comments. Since consumers do read text messages about products and sellers (Godes and Mayzlin 2004, Pavlou and Dimoka 2006), future research could examine the nature and role of product text comments on product quality perceptions and future product sales.

Finally, despite our attempt to overcome the under-reporting bias, it is interesting to examine what motivates consumers to write product reviews beyond bragging or moaning. For example, Dellarocas and Narayan (2006) proposed a statistical measure to predict whether consumers would write a product review. Future research could identify additional antecedents of the propensity to write product reviews across consumers.

CONCLUSION

Online WOM communication in the form of product reviews has become a major informational source for consumers and marketers, and the literature has used the mean of online product reviews to predict a product's quality and future success. However, since virtually all online product reviews have an asymmetric bimodal, positively-skewed (J-shaped) distribution due to two self-selection biases (purchasing and under-reporting), the mean of the online product reviews is shown to be a biased estimator of product quality. Under-reporting bias gives rise to a bimodal (U-shaped) distribution, and purchasing bias gives rise to an asymmetric, positively-skewed (J-shaped) distribution. To overcome these biases, this study develops the analytical 'brag-or-moan' model to identify a set of distributional parameters - (1) the mean, (2) standard deviation, (3) and the two modes (lower and upper) of the distribution of online product reviews - (to account for under-reporting bias), and also (4) product price (to account for self-selection bias) - that create a superior model for predicting future product sales compared to single-point estimators. This study aims to entice future research to derive superior models to understand the nature and value of online WOM communication by taking into account potential biases.

REFERENCES

1. Admati, A. R. and Pfleiderer, P. (2004), "Broadcasting Opinions with an Overconfident Sender," *International Economic Review*, 45, 2, 467-498.

2. Aigner, D.J. (1971), Basic Econometrics. Englewood Cliffs, New Jersey: Prentice-Hall.

3. Aitkin, M. and Rubin, D. B. (1985) "Estimation and hypothesis testing in finite mixture models," *Journal of the Royal Statistical Society B*, 47, 1, 67-75.

4. Anderson, Erin W. (1998) "Customer Satisfaction and Word of Mouth," *Journal of Service*, 1, 1, 5-17.

5. Arndt, J. (1967), "The Role of Product-Related Conversations in the Diffusion of a New Product," *Journal of Marketing Research*, 33, 291-295.

6. Banerjee, A., (1993), "The economics of Rumours", *Review of Economic Studies*, 60, 309-27.

7. Banerjee A. and Fudenberg D. (2004) "Word-of-Mouth Learning," *Games and Economic Behavior*, 46, 1, 1-22.

8. Bass, F.M, (1969), "A New Product Growth Model for Consumer Durables," *Management Science*, 15, 5, 215-227.

9. Basuroy S., S. Chatterjee and S. A. Ravid, (2003), "How critical are critical reviews? The box office effects of film critics, star-power, and budgets," *Journal of Marketing*, 67, 4, 105-117.

10. Biyalogorsky, E., Gerstner, E., and Libai, B. (2001), "Customer Referral Management: Optimal Reward Programs," *Marketing Science*, 20, 1, 82-95.

11. Brown, J. J. and Reingen, P.H. (1987), "Social Ties and Word-of-Mouth Referral Behavior," *Journal of Consumer Research*, 14, 4, 350-362.

12. Buttle, F.A. (1998) "Word-of-Mouth: Understanding and Managing Referral Marketing," *Journal of Strategic Marketing*, 6, 241-254.

13. Brynjolfsson, Erik, Hu Y., and Smith, Michael (2003), "Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety," *Management Science*, 49, 11, 1580-1596.

14. Chakravarti, I.M., Laha, R.G. and Roy, J. (1967) *Handbook of Methods of Applied Statistics*, Volume I, John Wiley and Sons, pp. 392-394.

15. Chatterjee, P. (2001), "Online Reviews: Do Consumers Use Them?" *Advances in Consumer Research*, 28, 1, 129-133.

16. Chen, P-Y., Wu, S-Y., and Yoon, J. (2004), "The Impact of Online Recommendations and Consumer Feedback on Sales," *Proceedings of the International Conference on Information Systems*, Washington, D.C., 711-724.

17. Chen, Y. and Xie, J. (2005), "Third-party Product Review and Firm Marketing Strategy," *Marketing Science,* 24 (2), 218-240.

18. Cheng, M-Y. and Hall, P. (1998) "Calibrating the Excess Mass and Dip Tests of Modality," *Journal of the Royal Statistical Society B*, 60, 3, 579-589

19. Chevalier, J. and Goolsbee, A. (2003), "Measuring Prices and Price Competition Online: Amazon and Barnes and Noble," *Quantitative Marketing and Economics*, 1(2) 203-222.

20. Chevalier, J. and Mayzlin, Dina (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 43, 3.

21. Clemons, Erik K., Gao, Guodong, and Hitt, L.M. (2006), "When Online Reviews Meet Hyper Differentiation: A Study of Craft Beer Industry," *Journal of Management Information Systems*, 23, 2, 149-171.

22. Dellarocas, Chrysanthos (2003), "The Digitization of Word-of-Mouth: Promise and Challenges of Online Reputation Mechanisms," *Management Science*, 49, 10, 1407-1424.

23. Dellarocas, Chrysanthos, Awad, Neveen, and Zhang, X. (2004), "Exploring the Value of Online Reviews to Organizations: Implications for Revenue Forecasting and Planning," *Proceedings of the 24th International Conference on Information Systems*, Washington D.C.

24. Dellarocas, Chrysanthos and Narayan, Ritu (2006), "A Statistical Measure of a Population's Propensity to Engage in Post-Purchase Online Word-of-Mouth," *Statistical Science*, 21, 2, 277-285.

25. Duan, W., Gu, Bin, and Whinston, Andrew (2005), "Do Online Reviews Matter? An Empirical Investigation of Panel Data," *Working Paper*, University of Texas at Austin.

26. Dalgaard, P. (2004), *Introductory Statistics with R,* Springer-Verlag, New York, NY.

27. Davidson R. and MacKinnon, J.G. (1981), "Several Tests for Model Specification in the Presence of Alternative Hypotheses," *Econometrica*, 49, 3, 781-793.

28. Davidson, R. and Mackinnon, J.G. (1993) *Estimation and Inference in Econometrics*, Oxford University Press, New York, NY.

29. Diebold, F.X., (2001) *Elements of Forecasting*, 2nd Edition., South Western Publishing, 2001

30. Eliashberg, J., Jonker J., Sawhney, M., and Wierenga, B. (2000) "MOVIEMOD: An Implementable Decision-Support System for Prerelease market Evaluation of Motion Pictures," *Marketing Science*, 19, 3, 226-243.

31. Eliashberg, J. and Shugan, S. (1997), "Film Critics: Influencers or Predictors?" *Journal of Marketing*, 61, 2, 68-78.

32. Ellison, G. and Fudenberg, D. (1995), "Word-of-Mouth Communication and Social Learning," *Quarterly Journal of Economics*, 110, 1, 93-125.

33. Everitt, B. S. and Hand, D. J. (1981), *Finite Mixture Distributions*, Chapman and Hall, London, England.

34. Fox, J (2002), *An R and S-Plus Companion to Applied Regression*, Sage Publications, Thousand Oaks, CA.

35. Ghose, Anindya, Michael Smith, and Rahul Telang (2006), "Internet Exchanges for Used Books: An Empirical Analysis of Product Cannibalization and Welfare Impact," *Information Systems Research* 17, 1, 3-19.

36. Godes, D. and Mayzlin, Dina (2004), "Using Online Conversations to Study Word of Mouth Communication," *Marketing Science*, 23, 4, 545–560.

37. Goldenberg J., B. Libai, S. Moldovan, E. Muller, (2004), "The Economic Implications of Negative Word of Mouth: A Dynamic Small-World Approach", *Working Paper*, Tel-Aviv University.

38. Hartigan, P. M. (1985), "Computation of the DIP Statistic to Test for Unimodality," *Applied Statistics,* 34, 320–325.

39. Hartigan, J. A. and Hartigan, P. M. (1985), "The DIP Test of Unimodality," *Annals of Statistics*, 13, 1, 70–84.

40. Herr, P. M., Kardes, F. R., and Kim, J. (1991) "Effects of Word-Of-Mouth and Product-Attitude Information on Persuasion: An Accessibility-Diagnosticity Perspective," *Journal of Consumer Research*, 17, 1, 454-462.

41. Henderson, D.J., Attar, P.K., and Russell, R.R. (2002), "Modality tests for use in Applied Econometrics: Application to Macroeconomic Convergence", *Working Paper*, University of California, Riverside.

42. Hoyer, W.D. and MacInnis, D.J. (2001), *Consumer Behavior,* (2nd Edition), Houghton Mifflin Company: Boston, MA.

43. Jensen A. (1969), "How much can we boost IQ and scholastic achievement?" *Harvard Educational Review*, 39, 1, 1-123.

44. Kadet, A. (2007), "Rah-Rah Ratings Online," *SmartMoney Magazine*, February 23, 2007, Online: http://www.smartmoney.com/toughcustomer/index.cfm?story=march2007&src=fb&nav=RSS20

45. Krider, R. and Weinberg, C. (1998), "Competitive Dynamics and the introduction of New Products: The Motion Picture Timing game," *Journal of Marketing Research*, 35, 1, 1-15.

46. Li, X. and Hitt, Lorin M. (2004), "Self selection and information role of online product reviews," *Workshop on Information Systems and Economics* (WISE 2004), Washington, DC.

47. Liu, Yong (2006), "Word-of-Mouth for Movies: Its Dynamics and Impact on Box Office Revenue," *Journal* of Marketing, 70, 3, 74-89.

48. Mahajan, V., Muller, E. and Kerin, R.A. (1984), "Introduction Strategy for new Products with Positive and Negative Word-of-Mouth," *Management Science*, 30, 12, 1389-1404.

49. Mizerski, R.W. (1982), "An Attribution Explanation of the Disproportionate Influence of Unfavorable Information," *Journal of Consumer Research*, 9, 4, 301-310.

50. Pareto, V. (1896), *Cours d'economie politique.* Rouge , Lausanne, France. 1964. G. H. Bousquet, G. Busino, (Eds.) Oevres Completes de Vilfredo Pareto, Volume 1, Librairie Droz, Geneva, Switzerland.

51. Pavlou, Paul A. and Angelika Dimoka (2006), "The Nature and Role of Feedback Text Comments in Online Marketplaces: Implications for Trust Building, Price Premiums, and Seller Differentiation," *Information Systems Research*, 17, 4, 391-412.

52. Ripley, B. D. (2001), "The **R** Project in Statistical Computing," *MSOR Connections: The Newsletter of the LTSN Mathematics, Statistics & Operations Research Network.*, 1, 1, 23-25

53. Reichheld, F. (2003), "The One Number You Need to Grow," *Harvard Business Review*, 81, 12, 46-54.

54. Richins, M.L. (1983), "Negative Word of Mouth by Dissatisfied Consumers: A Pilot Study," *Journal of Marketing*, 47, 1, 68-78.

55. Roeder, K. (1994), "A Graphical Technique for Determining the Number of Components in a Mixture of Normals," *Journal of the American Statistical Association*, 89, 487-495.

56. Schnapp, M. and Allwine, T. (2001), "Mining of Book Data from Amazon," SIMS Web Mining Conference, Online: http://www2.sims.berkeley.edu/resources/affiliates/workshops/webmining/Slides/ORA.ppt

57. Senecal, S. and Nantel, J., (2004), "The Influence of Online Product Recommendations on Consumers' Online Choices," *Journal of Retailing*, 80, 159–169.

58. Sheth, J.N., (1971), "Word of mouth in low risk innovations," *Journal of Advertising Research*, 11, 15–18.

59. Sorensen, A.T. and Rasmussen, S. (2004), "Is Any Publicity Good Publicity? A Note on the Impact of Book Reviews" *Working Paper*, University of Stanford.

60. Stephens, M. A. (1974), "EDF Statistics for Goodness of Fit and Some Comparisons," *Journal of the American Statistical Association*, 69, 730-737.

61. Thode Jr., H.C. (2002), *Testing for Normality*. Marcel Dekker, New York, NY.

62. Zipf, G. (1949), *Human Behavior and the Principle of Least Effort*. Addison-Wesley, Cambridge, MA.

Appendix 1: Procedures for Estimating the DIP Statistics

DIP is a measure of departure from unimodality. For N observations, the DIP statistic may be computed in order N operations. Hartigan and Hartigan (1985) show that the appropriate null distribution for the DIP test is uniform because DIP is asymptotically larger for the uniform distribution than for any distribution in a wide class of unimodal distributions. The authors determine the distribution of the DIP statistic by sampling from the space of uniform distributions. The basic idea is that, as the sample size grows, the DIP value for a unimodal distribution would approach zero, while the DIP of a multi-modal distribution would approach a positive constant.

A distribution function F is unimodal with mode m if F is convex in $(-\infty, a]$ and concave in $[a, \infty)$, that is function F increases in $(-\infty, a]$ and decreases in $[a, \infty)$. Hartigan and Hartigan (1985) define two more functions: the greatest convex minorant (g.c.m.) of F in $(-\infty, a]$ and the least concave majorant (l.c.m.) of F in $[a, \infty)$.

g.c.m = Max G(x) for x≤a

G represents all the functions that are convex in $(-\infty, a]$, and nowhere greater than F.

l.c.m= Min G(x) for x≥a,

G represents all the functions that are concave in $[a, \infty)$, and nowhere less than F.

The DIP algorithm is the following (Hartigan and Hartigan 1985, p. 79):

Let $x = x_1, x_2, \ldots, x_N$ be the atoms of F where the x's are listed in ascending order. To compute the minorant and majorant functions and the DIP statistic, the following steps must be executed:

(1) Begin with $x_L = x_1$, $x_U = x_N$, and $D = 0$.

(2) Compute the g.c.m., say G, and the l.c.m., say L, for F in $[x_L, x_U]$; denote the points of contact with F by g_1, g_2, ... g_K and l_1, l_2, … l_J, respectively.

(3) Suppose $d := \sup |G(g_i)-L(g_i)| > \sup |G(l_i)-L(l_i)|$ and that the supremum occurs at g_i satisfying $l_j \leq g_i \leq l_{j+1}$

Define $\chi_L^0 = g_i$ and $\chi_U^0 = l_{j+1}$

(4) Suppose $d := \sup |G(g_i)-L(g_i)| > \sup |G(l_i)-L(l_i)|$ and that the supremum occurs at g_j satisfying $g_j \leq l_j \leq gj+1$

Define $\chi_L^0 = g_i$ and $\chi_U^0 = l_j$

(5) If $d \leq D$, stop and set $D(F) = D$.

(6) If $d > D$, set:

$$D = \sup \left\{ D, \sup_{X_L \leq X \leq X_L^0} |G(x) - F(x)|, \sup_{X_U^0 \leq X \leq X_U} |L(x) - F(x)| \right\}$$

(7) Set $x_U = x_U^0$, $x_L = x_L^0$; return to (2)

Appendix 2: Proof of Propositions

Proof of Proposition 1:

When $p < \underline{\delta} < \overline{\delta}$, the distribution of reported quality $r_i = q + v_i$ is a normal distribution $N(q,\sigma^2)$ but with a window as described in Equation (3). The density function is derived as

$$\lambda(r_i) = \begin{cases} 0 & r_i < p \text{ and } \underline{\delta} < r_i < \overline{\delta}; \\[2ex] \dfrac{\phi(\frac{r_i - q}{\sigma})/\sigma}{(\Phi(\frac{\underline{\delta} - q}{\sigma}) - \Phi(\frac{p - q}{\sigma})) + (1 - \Phi(\frac{\overline{\delta} - q}{\sigma}))} & \text{otherwise.} \end{cases} \quad (A1)$$

The mean is given by

$$\mu_r = \int r_i \lambda(r_i) dr_i$$

$$= \int_p^{\underline{\delta}} \frac{r_i \phi(\frac{r_i - q}{\sigma})/\sigma}{(\Phi(\frac{\underline{\delta} - q}{\sigma}) - \Phi(\frac{p - q}{\sigma})) + (1 - \Phi(\frac{\overline{\delta} - q}{\sigma}))} dr_i + \int_{\overline{\delta}}^{+\infty} \frac{r_i \phi(\frac{r_i - q}{\sigma})/\sigma}{(\Phi(\frac{\underline{\delta} - q}{\sigma}) - \Phi(\frac{p - q}{\sigma})) + (1 - \Phi(\frac{\overline{\delta} - q}{\sigma}))} dr_i \quad (A2)$$

$$= q + \Lambda\sigma$$

where $\Lambda = \dfrac{\phi(\frac{\overline{\delta} - q}{\sigma}) - \phi(\frac{\underline{\delta} - q}{\sigma}) + \phi(\frac{p - q}{\sigma})}{(\Phi(\frac{\underline{\delta} - q}{\sigma}) - \Phi(\frac{p - q}{\sigma})) + (1 - \Phi(\frac{\overline{\delta} - q}{\sigma}))}$.

The variance is obtained by:

$$\sigma_r^2 = \int r_i^2 \lambda(r_i) dr_i - \mu_r^2$$
$$= \sigma^2 + \Lambda\sigma(\sigma - q) - \Lambda^2\sigma^2 \quad (A3)$$
$$= (1 + \Lambda - \Lambda^2)\sigma^2 - \Lambda q\sigma$$

Applying the Central Limit Theorem, we have

$$\lim_{n \to \infty} \Pr[\frac{n^{1/2}(\overline{r}_n - \mu_r)}{\sigma_r} \leq x] = \Phi(x) \quad (A4)$$

Consequently, the distribution of the average report score still converges to a normal distribution, except that the mean and variance are μ_r and σ_r^2 respectively now. Since $\mu_r = q + \Lambda\sigma$, the sample mean may be biased, and it will converge to true product quality q if and only if $\Lambda = 0$, which is true if and only if:

$$\phi(\frac{\overline{\delta} - q}{\sigma}) = \phi(\frac{\underline{\delta} - q}{\sigma}) - \phi(\frac{p - q}{\sigma}). \quad\quad\quad Q.E.D.$$

Proof of Proposition 2:

Plugging in Equation (5) to (6), we have $D(q) = 1 - \Phi(p - \bar{r}_n + \Lambda\sigma)$.

Since $\Lambda = \dfrac{\phi(\frac{\bar{\delta}-q}{\sigma}) - \phi(\frac{\underline{\delta}-q}{\sigma}) + \phi(\frac{p-q}{\sigma})}{(\Phi(\frac{\underline{\delta}-q}{\sigma}) - \Phi(\frac{p-q}{\sigma})) + (1 - \Phi(\frac{\bar{\delta}-q}{\sigma}))}$, we have:

$$\frac{\partial\Lambda}{\partial p} = \frac{\phi(\frac{p-q}{\sigma})[(\frac{q-p}{\sigma})((\Phi(\frac{\underline{\delta}-q}{\sigma}) - \Phi(\frac{p-q}{\sigma})) + (1 - \Phi(\frac{\bar{\delta}-q}{\sigma})) + 1]}{\sigma[(\Phi(\frac{\underline{\delta}-q}{\sigma}) - \Phi(\frac{p-q}{\sigma})) + (1 - \Phi(\frac{\bar{\delta}-q}{\sigma}))]^2} > 0;$$

$$\frac{\partial\Lambda}{\partial\underline{\delta}} = \frac{\phi(\frac{\underline{\delta}-q}{\sigma})[\phi(\frac{\underline{\delta}-q}{\sigma}) - \phi(\frac{\bar{\delta}-q}{\sigma}) - (\frac{q-\underline{\delta}}{\sigma})((\Phi(\frac{\underline{\delta}-q}{\sigma}) - \Phi(\frac{p-q}{\sigma})) + (1 - \Phi(\frac{\bar{\delta}-q}{\sigma})))]}{\sigma[(\Phi(\frac{\underline{\delta}-q}{\sigma}) - \Phi(\frac{p-q}{\sigma})) + (1 - \Phi(\frac{\bar{\delta}-q}{\sigma}))]^2};$$

and

$$\frac{\partial\Lambda}{\partial\bar{\delta}} = \frac{\phi(\frac{\bar{\delta}-q}{\sigma})[\phi(\frac{\bar{\delta}-q}{\sigma}) - \phi(\frac{\underline{\delta}-q}{\sigma}) - (\frac{\bar{\delta}-q}{\sigma})((\Phi(\frac{\underline{\delta}-q}{\sigma}) - \Phi(\frac{p-q}{\sigma})) + (1 - \Phi(\frac{\bar{\delta}-q}{\sigma})))]}{\sigma[(\Phi(\frac{\underline{\delta}-q}{\sigma}) - \Phi(\frac{p-q}{\sigma})) + (1 - \Phi(\frac{\bar{\delta}-q}{\sigma}))]^2}.$$

Then, we have the following results:

(i)
$$\frac{\partial D(q)}{\partial\bar{r}_n} = \phi(p - \bar{r}_n + \Lambda\sigma) > 0;$$

$$\frac{\partial D(q)}{\partial p} = -\phi(p - \bar{r}_n + \Lambda\sigma)(1 + \sigma\frac{\partial\Lambda}{\partial p}) < 0.$$

(ii) When both $\bar{\delta} - q$ and $q - \underline{\delta}$ are not too small, or more precisely, when

$$\phi(\frac{\bar{\delta}-q}{\sigma}) < \phi(\frac{\underline{\delta}-q}{\sigma}) + (\frac{\bar{\delta}-q}{\sigma})((\Phi(\frac{\underline{\delta}-q}{\sigma}) - \Phi(\frac{p-q}{\sigma})) + (1 - \Phi(\frac{\bar{\delta}-q}{\sigma})))$$

and

$$\phi(\frac{\underline{\delta}-q}{\sigma}) < \phi(\frac{\bar{\delta}-q}{\sigma}) + (\frac{q-\underline{\delta}}{\sigma})((\Phi(\frac{\underline{\delta}-q}{\sigma}) - \Phi(\frac{p-q}{\sigma})) + (1 - \Phi(\frac{\bar{\delta}-q}{\sigma}))),$$

we have $\dfrac{\partial\Lambda}{\partial\bar{\delta}} < 0$ and $\dfrac{\partial\Lambda}{\partial\underline{\delta}} < 0$.

Therefore,

$$\frac{\partial D(q)}{\partial\bar{\delta}} = -\sigma\phi(p - \bar{r}_n + \Lambda\sigma)\frac{\partial\Lambda}{\partial\bar{\delta}} > 0;$$

$$\frac{\partial D(q)}{\partial \underline{\delta}} = \sigma\phi(p - \bar{r}_n + \Lambda\sigma)\frac{\partial \Lambda}{\partial \underline{\delta}} < 0.$$

(iii) $\frac{\partial D(q)}{\partial \sigma} = -\phi(p - \bar{r}_n + \Lambda\sigma)(\Lambda + \frac{\partial \Lambda}{\partial \sigma}\sigma)$, therefore the sign depends on:

$$\Lambda + \frac{\partial \Lambda}{\partial \sigma}\sigma = \frac{(1 + (\frac{\bar{\delta} - q}{\sigma})^2)\phi(\frac{\bar{\delta} - q}{\sigma}) - (1 + (\frac{\underline{\delta} - q}{\sigma})^2)\phi(\frac{\underline{\delta} - q}{\sigma}) + (1 + (\frac{p - q}{\sigma})^2)\phi(\frac{p - q}{\sigma})}{(\Phi(\frac{\underline{\delta} - q}{\sigma}) - \Phi(\frac{p - q}{\sigma}) + 1 - \Phi(\frac{\bar{\delta} - q}{\sigma}))} -$$

$$\frac{(\frac{\bar{\delta} - q}{\sigma}\phi(\frac{\bar{\delta} - q}{\sigma}) - \frac{\underline{\delta} - q}{\sigma}\phi(\frac{\underline{\delta} - q}{\sigma}) + \frac{p - q}{\sigma}\phi(\frac{p - q}{\sigma}))(\phi(\frac{\bar{\delta} - q}{\sigma}) - \phi(\frac{\underline{\delta} - q}{\sigma}) + \phi(\frac{p - q}{\sigma}))}{[(\Phi(\frac{\underline{\delta} - q}{\sigma}) - \Phi(\frac{p - q}{\sigma})) + (1 - \Phi(\frac{\bar{\delta} - q}{\sigma}))]^2},$$

which could be either positive or negative, depending on the degrees of purchasing and under-reporting bias.

By the second equation of (5), we have:

$$\frac{\partial \sigma_r}{\partial \sigma} = (1 + \Lambda - \Lambda^2)\frac{\sigma}{\sigma_r} + \frac{\sigma^2}{2\sigma_r}\frac{\partial \Lambda}{\partial \sigma}(1 - 2\Lambda) - \frac{q}{2\sigma_r}(\Lambda + \sigma\frac{\partial \Lambda}{\partial \sigma})$$

whose sign is also undetermined.

Since

$$\frac{\partial D(q)}{\partial \sigma_r} = \frac{\partial D(q)}{\partial \sigma}\frac{\partial \sigma}{\partial \sigma_r},$$

the standard deviation of the online product reviews σ_r has an undetermined effect on product demand depending on the degrees of the purchasing bias and the under-reporting bias. *Q.E.D.*