Singapore Management University

Institutional Knowledge at Singapore Management University

Research Collection School Of Computing and Information Systems | School of Computing and Information Systems

6-2017

On self-selection biases in online product reviews

Nan HU
Singapore Management University, nanhu@smu.edu.sg

Paul A. PAVLOU

Jie ZHANG

Follow this and additional works at: https://ink.library.smu.edu.sg/sis_research

Part of the Databases and Information Systems Commons, E-Commerce Commons, and the Numerical Analysis and Scientific Computing Commons

Citation

HU, Nan; PAVLOU, Paul A.; and ZHANG, Jie. On self-selection biases in online product reviews. (2017). *MIS Quarterly*. 41, (2), 449-472.
Available at: https://ink.library.smu.edu.sg/sis_research/8012

This Journal Article is brought to you for free and open access by the School of Computing and Information Systems at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection School Of Computing and Information Systems by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email cherylds@smu.edu.sg.

MIS Quarterly

RESEARCH ARTICLE

ON SELF-SELECTION BIASES IN ONLINE PRODUCT REVIEWS[1]

Nan Hu
School of Business, Stevens Institute of Technology, 1 Castle Point Terrace,
Hoboken, NJ 07030 U.S.A. {nhu4@stevens.edu}

Paul A. Pavlou
Fox School of Business, Temple University, 1801 N. Broad Street,
Philadelphia, PA 19122 U.S.A. {pavlou@temple.edu}

Jie Zhang
College of Business, University of Texas, Arlington, 701 S. Nedderman Drive,
Arlington, TX 76019 U.S.A. {jiezhang@uta.edu}

Online product reviews help consumers infer product quality, and the mean (average) rating is often used as a proxy for product quality. However, two self-selection biases, acquisition bias (mostly consumers with a favorable predisposition acquire a product and hence write a product review) and underreporting bias (consumers with extreme, either positive or negative, ratings are more likely to write reviews than consumers with moderate product ratings), render the mean rating a biased estimator of product quality, and they result in the well-known J-shaped (positively skewed, asymmetric, bimodal) distribution of online product reviews. To better understand the nature and consequences of these two self-selection biases, we analytically model and empirically investigate how these two biases originate from consumers' purchasing and reviewing decisions, how these decisions shape the distribution of online product reviews over time, and how they affect the firm's product pricing strategy. Our empirical results reveal that consumers do realize both self-selection biases and attempt to correct for them by using other distributional parameters of online reviews, besides the mean rating. However, consumers cannot fully account for these two self-selection biases because of bounded rationality. We also find that firms can strategically respond to these self-selection biases by adjusting their prices. Still, since consumers cannot fully correct for these two self-selection biases, product demand, the firm's profit, and consumer surplus may all suffer from the two self-selection biases. This paper has implications for consumers to leverage online product reviews to infer true product quality, for commercial websites to improve the design of their online product review systems, and for product manufacturers to predict the success of their products.

Keywords: Online product reviews, self-selection biases, product uncertainty, product quality, product value, consumer behavior, electronic commerce, analytical modeling, econometric models, sales forecasting

[1]Kartik Hosanagar was the accepting senior editor for this paper. Guodong Gao served as the associate editor.

The authors contributed equally to the manuscript and the author names are listed alphabetically.

The appendices for this paper are located in the "Online Supplements" section of the *MIS Quarterly*'s website (http://www.misq.org).

Introduction

The Internet has enhanced the scale and scope of user-generated product reviews by expanding far beyond traditional settings and reaching a virtually infinite number of consumers. A prevalent view is that consumers rely on online product reviews to infer product quality and make purchasing decisions (e.g., Chatterjee 2001; Chen and Xie 2005; Chevalier and Mayzlin 2006; Forman et al. 2008; Moe and Schweidel 2012). However, the voluntary nature of online product reviews does not result in *all* relevant consumers writing reviews, creating self-selection biases in who writes online product reviews, making it difficult to infer "true" product quality. True product quality (or "value," the difference between quality and price) is an abstract and subjective concept, particularly for products when heterogeneous taste is involved, such as music and movies. True product quality is herein defined as the aggregate assessment of all relevant consumers who may acquire and review a product.

The literature showed the existence of a positively skewed, asymmetric, bimodal (or J-shaped) distribution of online product reviews (e.g., Gao et al. 2015; Hu et al. 2006, 2009). We formally examine the nature and role of two self-selection biases that determine the J-shaped distribution:[2] first, *acquisition bias*—consumers with a favorable predisposition toward a product are more likely to purchase a product and write a review; second, *underreporting bias*—consumers with extreme ratings (positive or negative) are more likely to report their reviews than consumers with moderate ratings. In this study, using a combination of empirical, experimental, analytical, and simulation methodologies, we seek to answer the following questions:

1. How do the proposed self-selection biases affect consumers' ability to infer true product quality from online product reviews to make their subsequent purchasing decisions?

2. Do consumers realize the existence of these two self-selection biases, and do they attempt to correct for them? Are consumers successful in correcting for these two self-selection biases?

3. How do firms strategically adjust their pricing in response to these two self-selection biases to influence current and future consumers' product quality perceptions, posted ratings, and purchases?

To capture the nature and effects of these two self-selection biases, we built a two-period analytical model to answer our research questions. In each period, the firm decides a price, and then consumers make two sequential decisions: whether to purchase a product, and then whether to write a review based on their realized value after purchasing the product. Since price is determined by the firm, a consumer considers whether to purchase the product based on existing reviews, price, product attributes, and personal preferences. Because only consumers with positive expected net utility will acquire the product and have the opportunity to write a review, this is proposed to give rise to *acquisition bias*. After acquiring the product and experiencing its actual quality, consumers voluntarily write reviews depending on their satisfaction or dissatisfaction levels, which is proposed to give rise to *underreporting bias*. Our analytical model shows that the consumer's expected product rating deviates from true product quality because of both self-selection (acquisition bias and underreporting) biases. Hence, the two self-selection biases are very likely to render the mean rating a biased proxy of product quality.

The second-period consumers read the reviews posted by first-period consumers to form quality expectations and make their purchase decisions. To understand how consumers interpret reviews, we empirically examined whether consumers realize the proposed self-selection biases in online product reviews and whether they can correct for these biases. A sales forecasting model with empirical data from Amazon showed that consumers do realize the existence of both self-selection biases to a certain extent and attempt to correct them by including additional parameters beyond the mean rating. Accordingly, we examine to what extent consumers can correct for the proposed self-selection biases by examining two competing views: (1) rational consumers fully correct for self-selection biases in online product reviews and infer unbiased product quality from (biased) online product reviews; (2) boundedly rational consumers cannot fully correct for the self-selection biases in reviews to infer true product quality. Our empirical results support the second assumption that consumers are **not** fully rational, and they cannot fully account for the self-selection biases.

Given that consumers cannot overcome these two self-selection biases to infer true product quality from online product reviews, the firm can influence consumers' purchasing and reviewing behaviors by pricing its products strategically. Price not only directly affects consumer utility as a cost, but

[2]There may be other biases in online product reviews and other reasons for the observed J-shaped distribution. However, we offer evidence that these two self-selection biases can reasonably (but not exhaustively) explain the J-shaped distribution (Appendix A). We also conducted a field study that attributes the J-shaped distribution largely to the two self-selection biases (Appendix B). Based on these findings, in this paper, we focus on the nature and effects of these two self-selection biases and not on identifying an exhaustive list of all factors that explain the J-shaped distribution of online product reviews.

it also indirectly affects a consumer's utility through the chains of impacts on current consumers' review probability and rating scores, and then on future consumers' product quality expectation. We built a two-period dynamic model and derived analytical solutions to the firm's optimal pricing decisions over two periods by maximizing its second-period profits and total profits, respectively. Our numerical results suggest that firms can strategically respond to both consumer self-selection biases by strategically adjusting their prices. For example, when consumers have low prior quality expectations, or when customers are more likely to report negative reviews, the firm should lower its first-period prices. This will attract more consumers to buy and subsequently write more positive reviews, which will raise the second-period consumers' product quality expectation. Then the firm can charge a higher price in the second period to increase its profits. Still, since consumers cannot fully correct for the self-selection biases, product demand, the firm's profit, and consumer surplus may suffer from the two self-selection biases.

Figure 1 shows the study's research roadmap.

This study makes the following contributions: First, besides analytically supporting the well-documented existence of the J-shaped distribution of online product reviews, we also experimentally supported the determinants of the J-shaped distribution: purchasing (acquisition bias) and reporting (underreporting bias). While most previous studies mainly focused on a single type of self-selection bias, (e.g., Dellarocas and Narayan [2006] focused on underreporting bias and Li and Hitt [2008, 2010] on acquisition bias, and only one recent paper, Jiang and Guo [2015], considered both self-selection biases), we examined the role of both self-selection biases using a lab experiment, thus extending the literature that has largely relied on archival data (Table 1).

Second, we built a dynamic analytical model to study the consequences of these two self-selection biases on consumer purchasing decisions and the firm's pricing over time. We analytically derived and empirically verified whether consumers realize and act upon the self-selection biases. Understanding whether consumers are rational in terms of online product reviews is important for studying the firm's dynamic pricing strategy. Prior research on self-selection biases in online product reviews focused on how price affects the mean rating (e.g., Li and Hitt 2010) or how ratings change over time (Moe and Schweidel 2012; Moe and Trusov 2011) without examining whether consumers realize the existence of these biases. Our empirical results show that consumers are boundedly rational, and they cannot perfectly account for the self-selection biases, which can be used by firms to inform their pricing.

Third, we developed an analytical model to quantitatively analyze the effects of these two self-selection biases on the firm's pricing decisions. We found that the firm can strategically respond to these self-selection biases by adjusting its prices over time to affect consumer purchases and reviews, thus shaping product demand and profits. In sum, our analytical work contributes to the emerging literature on online products in the Information Systems and marketing literatures (Table 1) by providing a comprehensive and integrated approach to an increasingly important practical problem faced by consumers and firms due to self-selection biases in online product reviews.

The paper proceeds as follows: In the next section, we review the literature on the nature of online product reviews. In the subsequent section, we analytically model the sources of self-selection biases from consumers' decisions to purchase and review products, whether and how consumers recognize and attempt to correct for the self-selection biases, and the firm's pricing decision. Finally, we discuss the paper's contributions and implications for theory and practice.

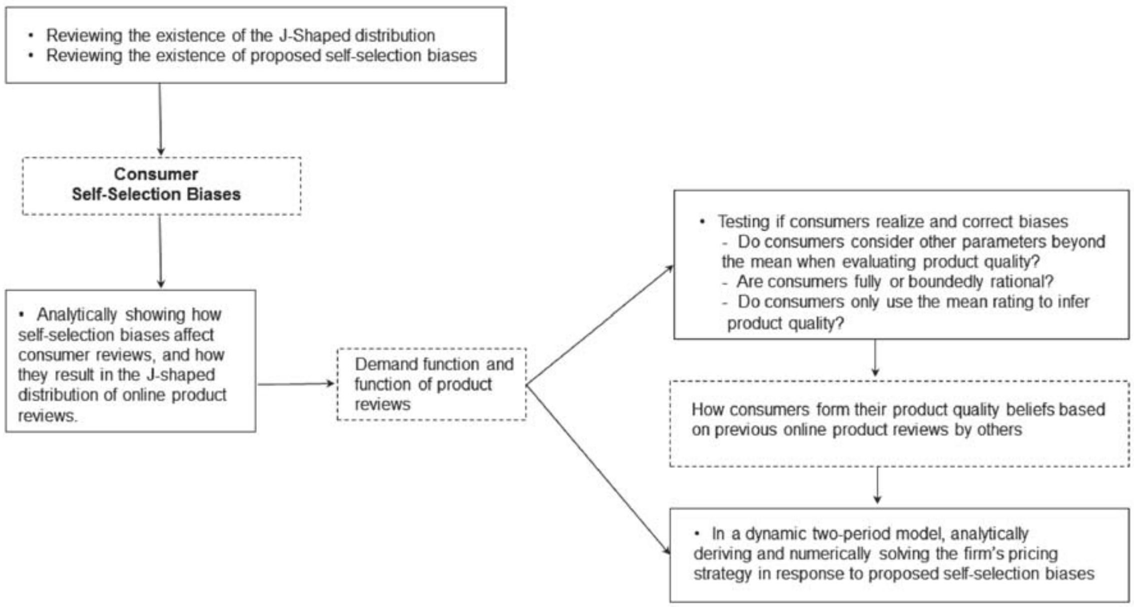
Literature Review: Online Product Reviews and Self-Selection Biases

The J-Shaped Distribution of Online Product Reviews

On most online product reviews sites, consumers can report an integer value on a five-point Likert-type scale, typically anchored at 1 star = least satisfied and 5 stars = most satisfied. The literature has shown the distribution of online product reviews to be positively skewed, asymmetric, bimodal (or J-shaped) (e.g., Gao et al. 2015; Hu et al. 2006, 2009). Appendix A specifies the existence of the J-shaped distribution of online product reviews for numerous products across multiple categories and several commercial websites (e.g., books, movies, apparel from Amazon, software from Download.com, and videos from Youtube.com), various mean ratings (e.g., 3-star, 3.5-star, and 4-star ratings), different phases of online product reviews over time, and online product reviews of different volume, such as more than 2,000 reviews and fewer than 20 reviews.

Self-Selection Biases in Online Product Reviews

Hu et al. (2009) proposed that two self-selection biases give rise to the observed J-shaped distribution: *acquisition bias* (or purchasing bias), where only consumers with a favorable pre-



Note: The dotted boxes represent the output from the upper level analyses, which are used by the lower level analysis.

Figure 1. Schematic Diagram of the Study's Research Roadmap

Table 1. Literature Review of Self-Selection Biases in Online Product Reviews

Papers	Acquisition Bias Considered	Under Reporting Bias Considered	Experimental Validation of the Two Self-selection Bias	Analytical Model of Acquisition Bias	Analytical Model of Under Reporting Bias	Impact of Acquisition Bias on Firm's Pricing Decision	Impact of Under-Reporting Bias on Firm's Pricing Decision	Validation of Whether Consumers Realize and Correct Biases
Dellarocas and Narayan (2006)		X						
Hu et al. (2006)		X			X			
Li and Hitt (2008)	X			X		X		
Hu et al. (2009)	X	X	X					
Li and Hitt (2010)	X			X		X		
Moe and Trusov (2011)	X							
Moe and Schweidel (2012)		X						
Jiang and Guo (2015)	X	X		X	X			
This paper	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**

Note: X represents whether the focal point was examined in the paper.

disposition toward a product would self-select to acquire a product, and thus have the opportunity to purchase and review the product, and *underreporting bias*, where consumers with either extreme levels of satisfaction or dissatisfaction with the product are more likely to self-select to report their reviews compared to consumers with moderate levels of satisfaction.

Acquisition Bias: Utility theory states that only consumers with an *ex ante* expected utility higher than or equal to the acquisition cost will acquire the product and have the opportunity to write a review. Pre-acquisition utility beliefs are formed by prior quality evaluations, WOM communication, or advertising (Yadav and Pavlou 2014). Acquisition bias exists because of the cost involved in acquiring a product, that is, the time and effort in searching for products, besides price. Consumers with heterogeneous expected utilities are separated into purchasers or non-purchasers, thus creating a self-selection bias in terms of who purchases and reviews the product. Li and Hitt (2008) identified a similar type of self-selection bias in online product reviews due to the different consumption timing chosen by each consumer. Utility theory implies that reviews tend to be written by consumers with higher utility (Kadet 2007), thus eliciting a positive self-selection bias in the distribution of online product reviews. For example, the mean of all Amazon's reviews in our data sample (Appendix A) is around 4.1 stars out of 5, consistent with Chevalier and Mayzlin (2006) who also showed book reviews on Amazon.com and BarnesandNoble.com to be mostly positive. This is also consistent with the endowment effect (Thaler 1980) that posits that consumers who already acquired a product tend to believe that the product has a higher value compared to those who did not. In contrast, consumers with low pre-acquisition utility are less likely to acquire the product, and they are less likely to write a (negative) review. Finally, cognitive dissonance theory predicts that the very fact of purchasing a product is associated with a higher consumer evaluation of the product.

Underreporting Bias: Underreporting bias exists since **not** all customers write online product reviews due to the time and effort needed. It is estimated that only 1 out of 1,000 people who purchase a product write a review on Amazon,[3] and only 1.6% of people write a comment after viewing a video on YouTube.com.[4] This can be explained by the satisfaction literature (Anderson 1998) which suggests a U-shaped relationship between satisfaction and WOM communication (e.g., Arndt 1967). Consumer satisfaction is generated by the difference of actual over expected utility (e.g., Anderson and Sullivan 1993). That is, when a consumer's realized utility greatly deviates from her prior expectations, she will be more motivated to exert the effort to "moan," to complain and warn others, or to "brag," to share the surprising good news and recommend the product to others.

In summary, these two self-selection biases help explain the J-shaped distribution of online product reviews (e.g., gao et al. 2015; Hu et al. 2006, 2009). Acquisition bias is responsible for the asymmetric right-skewness, while underreporting bias is responsible for the bimodality of the distribution. While underreporting bias can also create an asymmetric positive or negative skewness depending on whether consumers report more positive or negative reviews, respectively, the literature does not specify the valence of this skewness. Unless the positive and negative reviews perfectly cancel out, the self-selection biases would render the mean rating a biased estimator of product quality, and the mean rating would not reflect true product quality as rated by most relevant consumers, but rather a compromise between the high and low ratings.

The J-shaped distribution was also proposed to be caused by other factors besides the self-selection biases: First, the "true state of nature" may be that products are perceived as either outstanding or abysmal by consumers (Kadet 2007), and the J-shaped distribution simply represents the true perceptions of product quality across all consumers. This implies that all products would have a J-shaped (as opposed to a normal) distribution. However, it is quite unlikely that the true state of nature is that products fall into either extremely good or extremely bad quality with nothing in between. Second, consumers may be "overconfident" in their online product reviews (Admati and Pfleiderer 2004) and exaggerate their product assessments with either positive or negative ratings. This would suggest that all consumers would always post extreme ratings, which would only explain the deflation of the moderate ratings, but not the majority of positive ratings. Third, the extreme negative ratings may not be due to poor product quality, but due to poor service fulfillment, which is independent of the product. Analyses of the text content of online product reviews revealed that such reviews are very few (Hu et al. 2009), and unlikely to play a major role in shaping the J-shaped distribution. Finally, the extreme ratings could also be explained by paid reviewers who either attempt to promote or damage a product. However, since many products have a very large number of reviews, the extent of fraudulent manipulation is limited. The two latter explanations (service versus product fulfillment and paid reviewers) are unlikely to be prevalent across all products and time.[5]

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

[4]Youtube.com provides statistics of total views, which can be used to estimate the probability that a review is written.

[5]To test and rule out the first two likely explanations (true state of nature and consumer overconfidence), we conducted a simple lab experiment that compared the distribution of online reviews on Amazon versus a distribution of reviews from a random sample of consumers who was asked to sample and rate four products (music CD, movie DVD, software, textbook) (Appendix B). The results showed that the ratings of all potentially relevant consumers (subjects in the lab experiment) followed a unimodal normal distribution, implying that most consumers generally have moderate (and not extreme) views. Thus, the true state of nature is that product quality, when evaluated by virtually *all* relevant consumers, varies from low to high with the highest mass residing in the middle of the distribution. The lab experiment also showed that few subjects wrote extreme ratings, far fewer than those who wrote moderate ones. This negates the explanation of overconfident consumers. In sum, the lab experiment supports the basis for the observed J-shaped distribution due to the two self-selection biases.

Analytical Modeling of Self-Selection Biases in Online Product Reviews

The above empirical and experimental findings provide evidence for the existence of self-selection biases in online product reviews and how consumers' purchasing and reviewing behaviors are related to these biases. To further theoretically and formally examine how the proposed self-selection biases are formed, their effects on consumers, and how the firm responds to these self-selection biases, we build a two-period dynamic game model that formulates the consumers' and the firm's decision processes to derive equilibrium results.

As Figure 2 shows, in each period, a monopolist firm plays the Stackelberg leader by setting the price to maximize its expected profit including present and future, and a group of consumers enter the market and evaluate the product to form their prior quality beliefs. We distinguish between a product's search characteristics and experience characteristics. A consumer can directly observe search characteristics based on the product's online description, but she has to assess or infer the experience characteristics from other sources, such as online product reviews. Based on their resulting quality beliefs, consumers make a decision about whether to purchase a product based on their expected net utility. If a purchase is made, consumers will decide whether to review the product after experiencing the product, depending on their satisfaction levels. The reviews of the first-period consumers will shape their product quality expectations of the second-period consumers, and also directly influence the firm's second-period pricing, as we develop below.

In each period, we consider a three-stage decision problem: (1) the firm's pricing decision for a product, (2) the consumers' purchase decisions, and (3) the consumers' reviewing decisions. We used backward induction to solve the equilibrium in each period. Since we assume a group of consumers with the same distribution entering the market in each period, and each consumer only enters the market once and purchases no more than once, the consumers' decision is characterized the same across periods, except for different prior quality expectations and different prices. The firm's decisions, however, vary over time: it maximizes only the current period profit in the second period, while it maximizes the total profit of two periods in the first period. We first modeled the consumers' purchasing and reviewing decisions. The proposed self-selection biases influence consumers' quality expectations and their purchasing decisions, and their purchasing and reviewing decisions in turn render new product ratings, which further reshape consumers' self-selection biases. We also examined the impact of these self-selection biases on the distribution of online product reviews. We empirically examined consumers' awareness of these biases to verify how consumers form their prior quality expectations, then we solved the firm's optimal pricing decision in response to consumers' self-selection biases.

Consumers' Purchasing and Reviewing Decisions

Following Li and Hitt (2008), we define the net utility of consumer i by assessing a product's search and experience characteristics:

$$u_i = \theta_i + q_i - p \qquad [1]$$

where θ_i measures the consumer's heterogeneous preference toward the *observable* search attributes (attributes that can be inspected before purchase, such as title, manufacturer, brand, category, size, color, packaging, and description). The consumer knows the value of θ_i and price p before purchasing the product. Without loss of generality, we assume that a consumer only purchases at most one unit of the product. To make the model tractable, we assume that θ_i is normally distributed among the population of consumers with mean u_θ and variance σ_θ^2. q_i is consumer i's valuation of the product's *experience* characteristics (those that cannot be inspected prior to purchase, such as performance, reliability, touch, liking, and feeling). Consumer i does not know the value of q_i until she purchases and experiences the product. We assume that q_i is normally distributed with mean q and variance σ_q^2. Following Li and Hitt (2008), q represents intrinsic product quality (Milgrom and Roberts 1986), while standard deviation σ_q denotes product quality uncertainty. The coefficient of correlation between θ_i and q_i is denoted as ρ. When $\rho > 0$, a product with more favorable search attributes is more likely to result in a higher level of post-purchase quality, that is often more common. On the contrary, when $\rho < 0$, consumer i thinks a product with more favorable search attributes would be of lower quality.

Consumer i enters and inspects observable product attributes θ_i. Before purchasing the product and realizing quality q_i, consumers have a prior product quality expectation q^e. Besides other signals, consumers typically form their prior quality expectations based on reviews and other signals. Following Li and Hitt's (2010) argument, we assume that the first period quality expectation q_1^e is exogenous and common for all consumers. We will discuss how consumers in the second period form their quality expectation q_2^e later.

A consumer updates her *ex ante* quality expectation with the observed search attributes

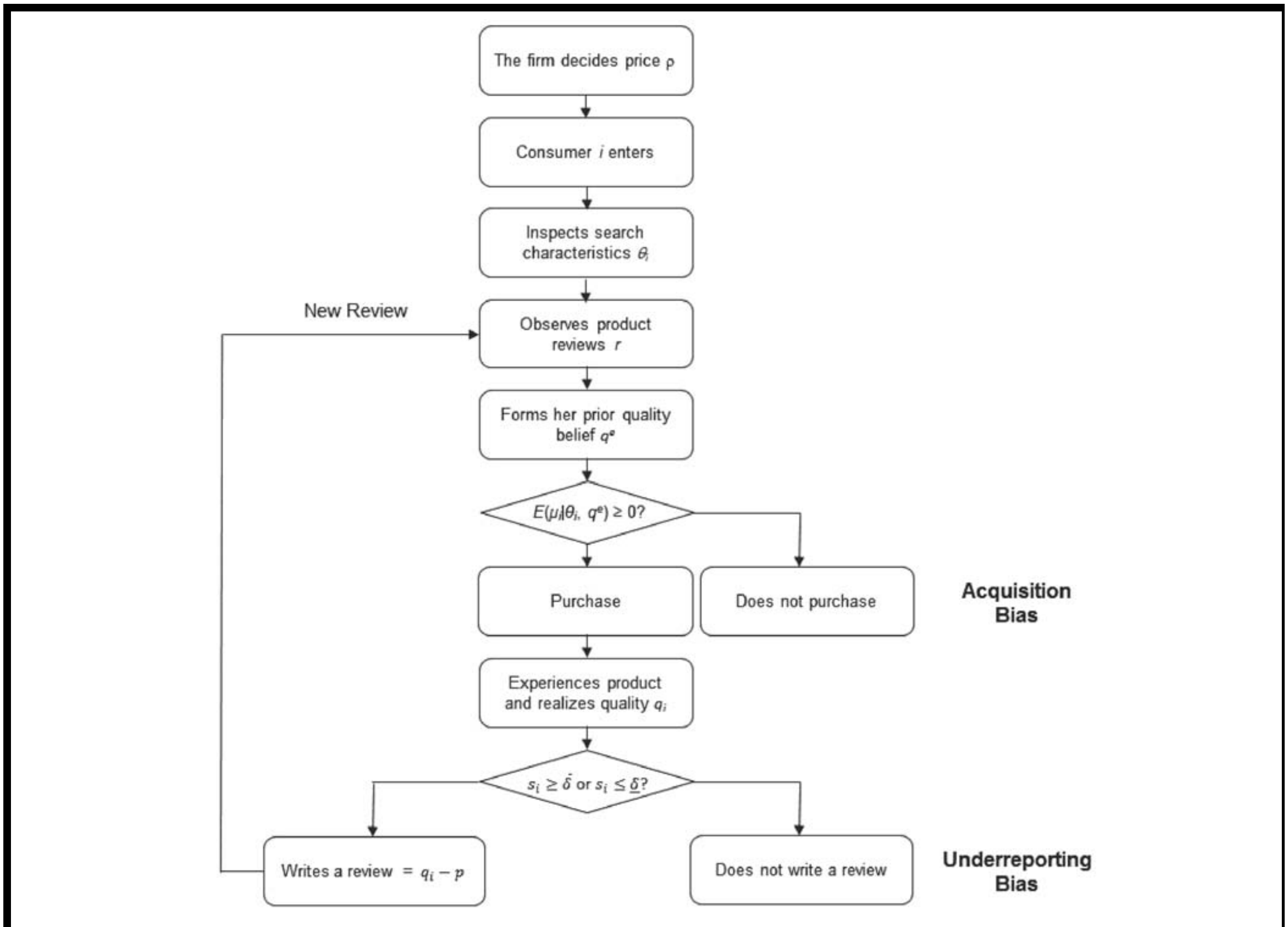


Figure 2. A Consumer's Purchasing and Reviewing Process and the Firm's Pricing Decision

$$E\left(q_i|\theta_i, q^e\right) = q^e + \rho\frac{\sigma_q}{\sigma_\theta}\left(\theta_i - \mu_\theta\right) \quad [2]$$

with conditional variance

$$Var\left(q_i|\theta_i\right) = \left(1-\rho^2\right)\sigma_q^2 \quad [3]$$

The consumer will purchase the product if and only if the expected net utility of purchasing the product is nonnegative, $E(u_i) \geq 0$. Combining Equations [1] and [2], we obtain the equivalent purchasing condition in which the consumer's preference for observable attributes θ_i has to exceed a threshold denoted as

$$\alpha\left(p, q^e\right) = \frac{\sigma_\theta\left(p - q^e\right) + \rho\sigma_q\mu_\theta}{\sigma_\theta + \rho\sigma_q} \quad [4]$$

Thus, only consumers with a θ_i greater than or equal to this threshold would acquire the product and have the opportunity to write a review, while consumers with $\theta_i < \alpha(p, q^e)$ are less predisposed toward the product, and they are thus screened out from writing a review. Thus, this purchasing self-selection is proposed to cause a self-selection bias in the reviews consumers write. We formally term this self-selection bias as ***acquisition bias***.

After acquiring and experiencing the product, consumer i realizes the product's experience attributes q_i and decides whether to write a review. Applying the satisfaction literature that assumes a U-shaped relationship between consumer satisfaction and WOM communication (e.g., Arndt 1967; Anderson 1998; Dellarocas and Narayan 2006), plus our experimental evidence on self-selection biases, we assume the incentive to write a review is mainly to *brag* or to *moan*. That is, consumers engage in WOM communication to express their strong satisfaction or dissatisfaction. When the experienced utility after purchase is close to her expectation, and thus the level of satisfaction or dissatisfaction is low, the consumer will have a low incentive to report a review. Following the satisfaction literature (Anderson and Sullivan 1993), satisfaction (or dissatisfaction), denoted by s_i, is assessed by the extent to which the consumer's *ex post* utility differs from her *ex ante* expected utility:

$$s_i = u_i - E(u_i|\theta_i) = (q_i - q^e) - \rho\frac{\sigma_q}{\sigma_\theta}(\theta_i - \mu_\theta) \quad [5]$$

A consumer is satisfied if her post-acquisition utility exceeds the prior expectation ($s_i > 0$), and is dissatisfied otherwise. To simplify the model, we assume the probability of consumer i reporting a product review to be given by a dichotomous reporting probability function [6]. This assumption is relaxed in [7].

$$Prob(report\ a\ rating) = \begin{cases} 1 & if\ s_i < \underline{\delta}\ or\ s_i > \overline{\delta} \\ 2 & otherwise \end{cases} \quad [6]$$

where $\underline{\delta}$ and $\overline{\delta}$ are the lower and upper bounds of a consumer's satisfaction within which she is not motivated to write a review. We assume $\underline{\delta}$ and $\overline{\delta}$ to be exogenous, and $\underline{\delta} \le \overline{\delta}$. Self-selection in reviewing a product is likely to under-represent the reviews of all relevant consumers. We term this self-selection bias ***underreporting bias***.

Following the literature (e.g., Li and Hitt 2010; Jiang and Guo 2015; Kuksov and Xie 2010), we assume that consumers rate the product based on their realized after-consumption net utility, $r_i = q_i - p$. When all consumers report their reviews without these two self-selection biases, the mean of ratings unbiasedly reflect the true value of the product, that is, $E(r_i) = q - p$. However, the (J-shaped) distribution of online product reviews is distorted by the two proposed self-selection biases, as demonstrated in the previous section with real-life empirical evidence.

Since the decision to write a product review is subject to the consumer's acquisition of the product, we use backward induction to solve the three-stage decision-making problem, starting from the reporting decision of a consumer who already acquired the product and after the firm's price decision. Considering the probability of a consumer reporting a product review (Equation [6]), the expected rating of consumer i is given in Lemma 1 (Appendix D provides proofs to the Lemma, Proposition, and Corollaries).

Lemma 1: For consumer i with preference toward the search attributes θ_i, prior quality expectation q^e who makes the purchase $E(u_i|\theta_i, q^e) \ge 0$, her expected rating can be expressed as

$$E(r_i|\theta_i, q^e) = q + \rho\tfrac{\sigma_q}{\sigma_\theta}(\theta_i - \mu_\theta) + \sqrt{1-\rho^2}\,\sigma_q\Lambda_0(q^e) - p$$

Underreporting bias is captured by $\sqrt{1-\rho^2}\,\sigma_q\Lambda_0(q^e)$ where

$$\Lambda_0(q^e) = \frac{\phi\left(\frac{\overline{\delta}(q^e-q)}{\sqrt{1-\rho^2}\sigma_q}\right) - \phi\left(\frac{\underline{\delta}+(q^e-q)}{\sqrt{1-\rho^2}\sigma_q}\right)}{\Phi\left(\frac{\underline{\delta}+(q^e-q)}{\sqrt{1-\rho^2}\sigma_q}\right) + \left(1-\Phi\left(\frac{\overline{\delta}+(q^e-q)}{\sqrt{1-\rho^2}\sigma_q}\right)\right)}$$

The condition for this bias alone to be equal to zero is either $\overline{\delta} = \underline{\delta} = 0$ or $\overline{\delta} + \underline{\delta} = 2(q - q^e)$. That is, the expected rating will not be distorted by underreporting bias when the consumer writes a review regardless of her satisfaction level, or when there is an equal number of satisfied and dissatisfied consumers who underreport. The impact of underreporting bias on the expected rating can be two-fold: when consumers are more likely to moan ($\overline{\delta} + \underline{\delta} > 2(q - q^e)$), fewer negative reviews will be underreported than positive ones, and the expected rating will underestimate actual product value; when consumers are more likely to brag ($\overline{\delta} + \underline{\delta} < 2(q - q^e)$), the expected rating will inflate actual product value.

For the second-stage decision, consumers make their purchasing decisions by comparing expected utility $E(u_i|\theta_i, q^e)$ and the outside option which is assumed to be 0. Therefore, consumers with a lower preference to search characteristics $\theta_i < \alpha(p, q^e)$ will refrain from purchasing the product and writing a product review. Thus, the distribution of the consumers' preference to search attributes θ_i is a truncated normal distribution on the support of $[\alpha(p, q^e), +\infty)$. Considering both self-selection biases, we derive the expected rating obtained from the group of consumers who make a decision to purchase a product in the first period in Proposition 1:

Proposition 1: Accounting for both acquisition and underreporting biases, a consumer reports a review with expected probability

$$\left(1-\Phi\left(\tfrac{p-q^e-\mu_\theta}{\sigma_\theta+\rho\sigma_q}\right)\right)\left(1-\left(\Phi\left(\frac{\overline{\delta}+(q^e-q)}{\sqrt{1-\rho^2}\sigma_q}\right)-\Phi\left(\frac{\underline{\delta}+(q^e-q)}{\sqrt{1-\rho^2}\sigma_q}\right)\right)\right)$$

The expected product rating can be expressed as

$$E(r) = q + \sqrt{1-\rho^2}\,\sigma_q\Lambda_0(q^e) + \rho\sigma_q\lambda(\alpha(p,q^e)) - p$$

and the variance of the rating σ_r^2 is expressed as

$$\rho^2\sigma_q^2\left(1-\lambda\left(\alpha\left(p,q^e\right)\right)\left(\lambda\left(\alpha\left(p,q^e\right)\right)-\alpha\left(p,q^e\right)\right)\right)$$
$$+\left(1-\rho^2\right)\sigma_q^2\left(1+\Lambda_1\left(q^e\right)-\Lambda_0^2\left(q^e\right)\right)$$

With probability $\Phi\left(\frac{p-q^e-\mu_\theta}{\sigma_\theta+\rho\sigma_q}\right)$, a consumer does not purchase a product, and thus does not write a product review, resulting in acquisition bias; with probability

$$\left(1-\Phi\left(\frac{p-q^e-\mu_\theta}{\sigma_\theta+\rho\sigma_q}\right)\right)\left(\Phi\left(\frac{\overline{\delta}+\left(q^e-q\right)}{\sqrt{1-\rho^2}\sigma_q}\right)-\Phi\left(\frac{\underline{\delta}+\left(q^e-q\right)}{\sqrt{1-\rho^2}\sigma_q}\right)\right),$$

she purchases the product but not report a review due to moderate (dis)satisfaction, resulting in underreporting bias. Otherwise, the consumer will write a review with expected rating: $E(r)=q+\sqrt{1-\rho^2}\sigma_q\Lambda_0\left(q^e\right)+\rho\sigma_q\lambda\left(\alpha\left(p,q^e\right)\right)-p$ where $\lambda(x)=\frac{\phi(x)}{1-\Phi(x)}$.

Proposition1 further reveals that the impact of acquisition bias on the expected rating is captured by $\rho\sigma_q\lambda\left(\alpha\left(p,q^e\right)\right)$. When the observable search attributes are positively correlated with product quality ($\rho > 0$), acquisition bias will inflate the expected product rating. Otherwise, the expected ratings will underestimate actual product quality. In the unique case of ρ being zero, the consumer's purchasing decision is totally random, and hence there is no systematic bias to the mean rating. Acquisition bias can be ignored when the acquisition cost is extremely low compared to the consumer's prior quality expectation $p << q^e-\rho\frac{\sigma_q}{\sigma_\theta}\mu_\theta$.

From Proposition 1, it is easy to show that the two bounds of review reporting $\underline{\delta}$ and δ have direct impact on the size of the underreporting bias, and further affect the consumers' expected rating.

Corollary 1: Given product price p and consumer quality expectation q^e, when $\overline{\delta}+\left(q^e-q\right)<\Lambda_0\left(q^e\right)\sqrt{1-\rho^2}\sigma_q$, the expected product rating decreases with $\underline{\delta}$ and increases with $\overline{\delta}$, vice versa.

When $\overline{\delta}+\left(q^e-q\right)<\Lambda_0\left(q^e\right)\sqrt{1-\rho^2}\sigma_q$, that is, when the bar for reporting positive ratings $\overline{\delta}$ is low and consumers are more likely to brag than moan (the underreporting bias $\Lambda_0\left(q^e\right)\sqrt{1-\rho^2}\sigma_q$ is relatively large), the majority of ratings reported are positive. In this case, a lower $\underline{\delta}$ reduces the amount of low value consumers from reporting potentially low ratings, thus it boosts the expected rating; while a higher $\overline{\delta}$ prevents some medium-high value consumers from reporting ratings, therefore also increasing the expected ratings. When $\overline{\delta}+\left(q^e-q\right)>\Lambda_0\left(q^e\right)\sqrt{1-\rho^2}\sigma_q$, that is, when the bar for reporting positive ratings $\overline{\delta}$ is high and consumers are more likely to moan than brag, negative reviews dominate. Thus, a higher $\underline{\delta}$ will add more medium-low ratings to dampen the extremely negative ratings, and increase the expected rating, while a lower $\overline{\delta}$ will include more high-value customers leaving positive feedback, and boost the expected rating.

To illustrate the effect of the two proposed self-selection biases in the distribution of online product reviews, we simulated the consumers' decision processes as described above. Given product price and other parameters[6] (p = 30, q^e = 100, $\underline{\delta}$ = -23; $\overline{\delta}$ = 64), we numerically generated 1,272 consumers entering the market, each with a random draw of θ_i and q_i from normal distribution with mean and variance μ_θ = 60, σ_θ = 25, q = 145, respectively. The correlation ρ bebeween θ_i and q_i is 0.1.[7] Consumers choose to purchase and review the product following the decision processes described in Figure 2. We normalized both the original and reported ratings by the *population* mean and variance of ratings, and we then aggregated them by rounding to the closest integer ratings (there are less than 2% of ratings are out of the bounds, which is negligible), such that it is numerically comparable to the 1 to 5 rating scale used in most review websites (Appendix A). We obtained the frequency charts of consumer reviews at different score levels. The resulting distribution of online product reviews with the consumers' self-selection biases is shown in Figure 3(b). In contrast, we also plotted the distribution of online product reviews for the whole population of customers in Figure 3(a). These figures clearly demonstrate that the proposed self-selection biases in writing reviews distorted the distribution from a normal to a J-shaped distribution, and shifted the real mean rating from 3.00 to the observed rating of 3.95.

We also relaxed the assumption about the dichotomous reporting rules with the probability function [6] to a general U-shaped probability function [7], which is defined as

[6]The number 23 represents the situation that consumers whose shock ranks between 2.5% and 75% (meaning that we rank consumers' shock from most negative to most positive) keep silent or are partially silent. In fact, if we fix X_L = 2.5% and change X_H from 75% to 99%, we can always observe a J-shaped distribution.

[7]To insure that the correlation is 0.1, we first generated two sequences of *independent* normal distributed random numbers X_1 and X_2. Then we defined a new sequence $X_3=\rho X_1+\sqrt{1-\rho^2}X_2$. This new X_3 sequence will have a correlation ρ with the X_1 sequence.

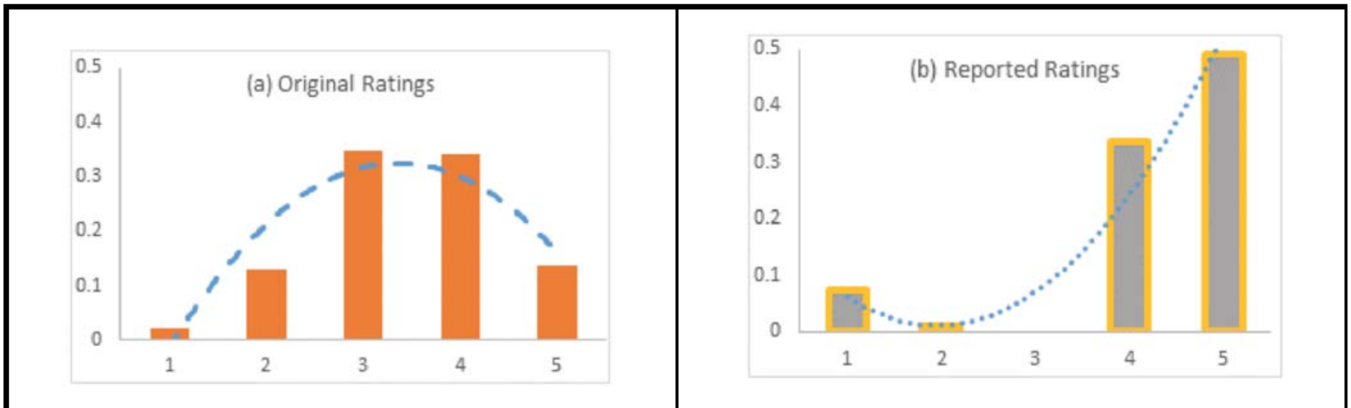


Figure 3. Simulated Distribution of Online Reviews (a) Without and (b) with Self-Selection Biases

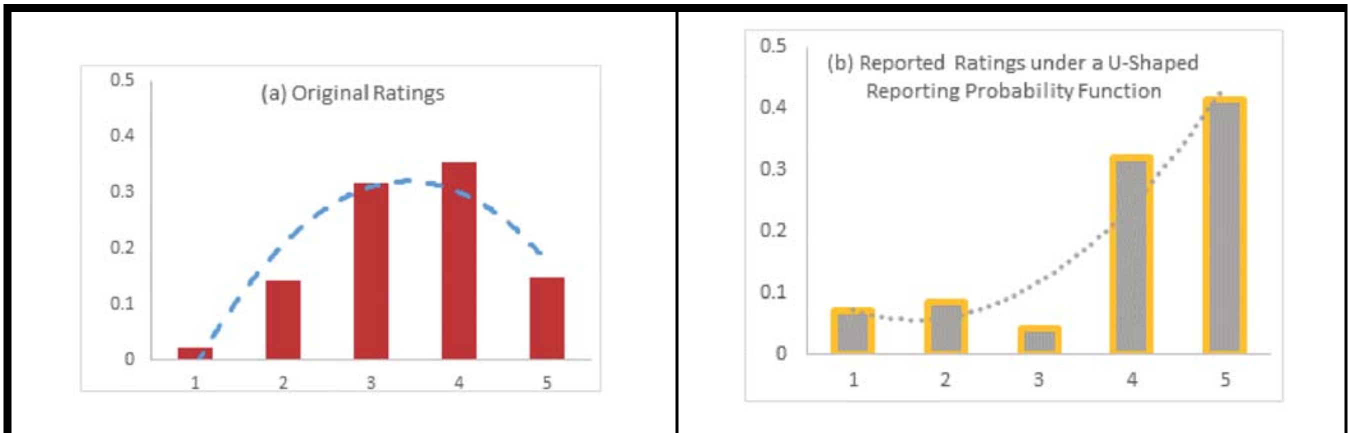


Figure 4. Simulated Distribution of Online Reviews (a) Without and (b) with Self-Selection Biases (with a U-Shaped Review Probability Function)

$$Prob(report\ a\ rating) = \begin{cases} \left(\frac{s_i - \min\{s_i\}}{\max\{s_i\} - \min\{s_i\}} - 0.5\right)^2 & if\ \underline{\delta} \le s_i \le \overline{\delta} \\ 1 & Otherwise \end{cases} \quad [7]$$

We first transformed the satisfaction (or dissatisfaction) level s_i to a range between 0 and 1. Then we chose 0.5, which is the center of that range and is also located between $\frac{\underline{\delta} - \min\{s_i\}}{\max\{s_i\} - \min\{s_i\}}$ and $\frac{\overline{\delta} - \min\{s_i\}}{\max\{s_i\} - \min\{s_i\}}$, as the lowest point of the U-shaped distribution function. The difference between the probability functions [6] and [7] is that function [6] assumes that consumers will not report reviews unless they are extremely satisfied or extremely unsatisfied, while function [7] allows consumers with a medium level of satisfaction to write reviews with a probability less than 1, and that probability increases when their satisfaction (or dissatisfaction) level s_i deviates from the center.

We simulated the distribution of online product reviews using the same parameter values as Figure 3, and we present the results in Figure 4(a) (without self-selection biases) and Figure 4(b) (with self-selection bias). Compared with Figure 3(b), Figure 4(b) shows a slightly higher density of ratings in the middle due to the relaxed assumption of the reporting probability function. However, we again observe a J-shaped distribution, while the original distribution of online product reviews from the whole population is normal (bell-shaped). Thus, even under a relaxed reporting rule, the two self-selection biases still not only distort the distribution of product reviews, but also make the mean rating unable to represent the true mass of consumer product utilities.

Lemma 1 suggests that even after taking out price, which is common for all consumers, the mean rating of online product reviews does *not* necessarily reflect true product value $(q - p)$

due to the existence of acquisition and underreporting biases. By Proposition 1, to remove the self-selection biases in the expected ratings requires $\sqrt{1-\rho^2}\sigma_q\Lambda_0\left(q^e\right)=-\rho^2\sigma_q\lambda\left(\alpha\left(p,q^e\right)\right)$. Thus, these self-selection biases will not distort the mean rating from unbiasedly reflecting product value except under the following very stringent conditions: (1) neither self-selection bias exists when all consumers acquire the product ($p \le q^e - \rho\frac{\sigma_q}{\sigma_\theta}\mu_\theta$), and all consumers write a review, or there is an equal number of satisfied and dissatisfied consumers who choose not to write an online product review ($\overline{\delta}$ = $\underline{\delta}$ = 0) or ($\overline{\delta}$ + $\underline{\delta}$ = $2(q - q^e)$); or (2) both self-selection biases exist and they exactly cancel each other out. However, there is a negligible probability for either condition to be satisfied in practice. Besides freeware and heavy discounted products, retailers are unlikely to set the price so low for *all* consumers to acquire the product; and there is a trivial probability that both biases would perfectly cancel each other out, as required by Condition (2). Therefore, taken together, the mean rating of online product reviews is likely to be biased because it is based on a truncated sample caused by the proposed self-selection biases. In that sense, consumers have to be cautious when using the mean rating to infer true product value. To accurately and reliably estimate true product value, we argue that consumers must go beyond the mean rating and consider the entire distribution of online product reviews.

Ignoring these self-selection biases may cause consumers to make wrong decisions in acquiring a product. The intuition is as follows: the mean rating includes both acquisition bias and underreporting bias (Proposition 1). Acquisition bias is positive when $\rho > 0$ and underreporting bias can be positive or negative depending on the consumer's intention to write a review (to brag or to moan) and her degree of (dis)satisfaction. Thus, the mean rating may be higher or lower than true value, depending on the net impact of these two self-selection biases. When the mean rating is inflated by the two self-selection biases, consumers who form their prior expectation about product value based on the mean rating will be misled, and consumers with a lower θ will purchase the product but realize a negative utility afterward. When the mean rating is deflated by the two self-selection biases, there will be some consumers who should have acquired the product and have enjoyed its utility, but they end up not purchasing the product because they are misled by the "low" mean rating. Therefore, if a consumer does not realize the two proposed self-selection biases in online product reviews, and she infers product value based solely on the mean rating, then the two proposed self-selection biases may hurt the consumer's utility.

Consumer Rationality on Self-Selection Biases and Evolution of Online Product Reviews

The analytical results presented in the previous subsection suggest that consumers should account for both acquisition bias and underreporting bias when relying on online product reviews to infer true product quality. This subsection empirically examines the consumers' awareness of these self-selection biases, and it also investigates how consumers draw their product quality expectations from online product reviews. In order to isolate the effect of the firm's pricing on online product reviews from the consumers' learning effect, we assumed price p as exogenous in the analyses throughout this subsection. We also chose the empirical data without significant price changes during the sample period.[8]

We first estimated a sales forecasting model to check consumer rationality regarding the two self-selection biases in online product reviews. Then we analyzed how different rationality assumptions affect consumers' prior quality expectations based on past reviews and the evolution of online product reviews: *rational* consumers should be able to realize and fully account for self-selection biases in online product reviews to infer true product value; *boundedly rational* consumers take existing online product reviews "as is" without correcting for self-selection biases. For each of these two scenarios, we derived analytical predictions of online product reviews generated over time by extending the model presented earlier, and we empirically tested the hypotheses to understand consumer rationality regarding the two self-selection biases and how consumers infer product value from online product reviews.

A Sales Forecasting Model

We develop a sales forecasting model to empirically test consumer awareness of the self-selection biases in online product reviews by testing whether consumers make purchase decisions purely based on the mean rating without considering other distributional parameters. If this is true, then we infer that consumers are completely unaware of the existence of self-selection biases. Otherwise, we can identify which distributional parameters of online product reviews consumers use to draw or correct for acquisition bias and underreporting bias.

Based on the analytical model, given a product with quality q, price p, and prior quality expectation q^e, the expected

[8]The price change variable at the 25% percentile, the 50% percentile, and the 75% percentile are all zero, and the mean percentage price change is only 0.63%.

product demand in a period can be derived through a consumer utility function (Equation [1]):

$$E\left(D\left(p,q^e\right)\right)=\int_{\alpha\left(p,q^e\right)}^{+\infty}dF_\theta(\theta)=1-\Phi\left(\frac{p-q^e-\mu_\theta}{\sigma_\theta+\rho\sigma_q}\right) \quad [8]$$

Given the expected demand function [8], we have

$$\frac{\partial E(D)}{\partial p}=-\frac{1}{\sigma_\theta+\rho\sigma_q}\phi\left(\frac{p-q^e-\mu_\theta}{\sigma_\theta+\rho\sigma_q}\right)<0$$

and

$$\frac{\partial E(D)}{\partial q^e}=\frac{1}{\sigma_\theta+\rho\sigma_q}\phi\left(\frac{p-q^e-\mu_\theta}{\sigma_\theta+\rho\sigma_q}\right)>0$$

That is, product demand decreases with price p, but it increases with prior quality expectation. The expected demand function [8] also suggests that $\frac{\partial E(D)}{\partial\sigma_\theta}=-\frac{q^e+\mu_\theta-p}{\left(\sigma_\theta+\rho\sigma_q\right)^2}\phi\left(\frac{p-q^e-\mu_\theta}{\sigma_\theta+\rho\sigma_q}\right)<0$, and $\frac{\partial E(D)}{\partial\sigma_q}=-\frac{\rho\left(q^e+\mu_\theta-\rho\right)}{\left(\sigma_\theta+\rho\sigma_q\right)^2}\phi\left(\frac{p-q^e-\mu_\theta}{\sigma_\theta+\rho\sigma_q}\right)<0$ given that ρ is generally positive. A higher variance of consumer preference toward observable attributes σ_θ^2 implies that consumers have diverse tastes, and a high quality variance σ_q^2 suggests higher quality uncertainty (e.g., Dimoka et al. 2012; Hong and Pavlou 2014). Both variances are negatively affecting product demand. Proposition 1 suggests that the variance of ratings σ_r^2 is positively correlated with σ_θ^2 and σ_q^2. Thus, we expect the standard deviation of online product reviews σ_r to be *negatively* related to future product sales (e.g., Clemons et al. 2006; Zhu and Zhang 2010).

We also want to verify whether consumers realize the existence of underreporting bias and attempt to correct for this bias by considering the two bounds $\overline{\delta}$ and $\underline{\delta}$ of the distribution of online product reviews in order to infer true product quality, which is presumably directly related to product sales. Because the values of the two theoretical bounds are not readily available, we use X_L and X_U returned by the DIP test (Hartigan and Hartigan 1985) as proxies for $\overline{\delta}$ and $\underline{\delta}$, respectively. Based on Hartigan and Hartigan, the DIP test measures multimodality in a sample by the maximum difference, over all sample points, between the empirical distribution function, and the best fitting unimodal distribution, which is the unimodal distribution function that minimizes that maximum difference. The dip of a distribution function measures the departure from unimodality. To find the dip d for an arbitrary distribution function F, there must exist a nondecreasing function G, for $X_L \le X_U$, where G is the greatest convex minorant of $F + d$ in $(-\infty, X_L)$, G has constant maximum slope in (X_L, X_U), and G is the least concave majorant in $F - d$ in $[X_U, \infty)$. The greatest convex minorant of F in $(-\infty, X_L)$ is sup $G(x)$ for $x \le a$, where the *sup* is taken over all functions G that are convex in $(-\infty, X_L)$ and nowhere greater than F; while the least concave minorant of F in $[X_U, \infty)$ is *inf* $L(x)$ for $x \ge a$, where the infinity is taken over all functions L, which are concave in $[X_U, \infty)$ and nowhere less than F.

X_L represents the point below which consumers will speak out, X_U represents the point above which consumers will speak out. By Corollary 1, when consumers are more likely to brag, as X_L increases, *ceteris paribus*, the mean rating of online product reviews will decrease, resulting in a decrease in sales. Also, under that scenario, when consumers are more likely to moan, as X_U increases, given everything else equal, the mean rating will increase, resulting in an increase in product sales. If the coefficients of the variables X_L and X_U are significant in predicting product sales, as we expect, then our conjecture would be supported, indicating that consumers do realize the existence of the underreporting bias to a certain degree. Otherwise, our conjecture about correcting for underreporting bias would be rejected.

To empirically test the predictive power of these parameters (product price, mean, volume of reviews, and the standard deviation of product ratings), and the two modes of the distribution of online product reviews, we developed a sales forecasting model with actual online product sales and reviews data from Amazon.com (Appendix A). Following other sales forecasting models in the literature (e.g., Clemons et al. 2006; Duan et al. 2008; Liu 2006), we use future product sales to measure product demand and use *SalesRank* as a proxy for product sales, which is *negatively* correlated with sales (Brynjolfsson et al. 2003; Chevalier and Goolsbee 2003; Ghose et al. 2006). The following log-linear regression model was used to predict product sales rank:[9]

[9]We collected our data from Amazon Web Service (AWS), and we constructed two datasets to examine our questions. The first dataset is cross-sectional data composed of a random sample of books, DVDs, and videos. For this dataset, we collected the product information and corresponding consumer reviews from Amazon.com in July 2005. The second dataset is a panel dataset composed of a sequence of features of online reviews (price, sales, and review information) for a sample of books, DVDs, and videos collected over several months at approximately three-day intervals. The initial set of products in this panel dataset was randomly chosen from Amazon in July 2005. For the panel data collection, since it occurs approximately every three days, we identified each data collection batch by a unique sequence number. Because we needed to know the historical sale, review, and price information, we used the panel dataset to answer the questions as to whether consumers are aware and correct for the self-selection bias. For the remaining research questions, we used the cross-sectional datasets.

Table 2. Regression on Future Product Sales (Dependent Variable: ln($SalesRank_{t+1}$))

Variable	Coefficient
$AvgRating_t$	-0.0513***
X_{Lt}	0.0573***
X_{Ut}	-0.0685***
$StdevRating_t$	0.0086*
ln($SalesRank_t$)	0.7114***
ln($Price_t$)	0.0780***
ln($NumRev_t$)	-0.0869***
Book_Dummy	0.311***
DVD_Dummy	-0.154***
Intercept	3.7890***
Adjusted R^2	77.29%

Note: ***p < .001; **p < .01; *p < .05; ^+p < .10

$$\ln(SalesRank_{t+1}) = \beta_0 + \beta_1 AvgRating_t + \beta_2 X_{Lt} + \beta_3 X_{Ut} + \beta_4 StdevRating_t + \beta_5 \ln(SalesRank_t) + \beta_6 \ln(Price_t) + \beta_7 \ln(NumRev_t) + \beta_8 Book\_Dummy + \beta_9 DVD\_Dummy + \varepsilon_1 \quad [9]$$

As shown in Table 2, controlling for prior sales rank and the number of online product reviews, the proposed parameters predict future product sales as expected: the *mean* of online product reviews ($\beta_1 = -0.0513, p < 0.01$), the two *modes* X_L ($\beta_2 = 0.0573$, $p < 0.01$) and X_U ($\beta_3 = -0.0685$, $p < 0.01$) and the *standard deviation* of online product reviews ($\beta_4 = 0.0086$, $p < 0.05$) are all significant in predicting sales rank.

These results suggest that consumers are not completely unaware of the existence of self-selection biases, and they do attempt to overcome these biases by going beyond the mean rating and by using other parameters, specifically by trying to infer the association between standard deviation, the two modes of the distribution of online product reviews, and product quality. These results also specify the particular distribution parameters that consumers rely upon, which denote the variation of the product ratings around the mean rating. Furthermore, as elaborated in Appendix C, the proposed self-selection controlled model explains at least 2% higher variance compared to the five competing models in terms of predicting future product sales, supporting our logic.

The Model for Rational Consumers Fully Correcting for Self-Selection Biases

Provided that the sales forecasting test revealed that consumers consider additional parameters related to the self-selection biases beyond the mean rating to evaluate product value, we now examine to what extent consumers have realized the self-selection biases, and to what extent they can correct them. In this subsection, we consider an extreme scenario, that is, consumers can rationally and fully discern the proposed self-selection biases in online product reviews, and can mentally separate self-selection biases to infer true product quality.

We expand the single-period consumer's purchasing and reviewing decisions described earlier into a dynamic problem of sequentially entering consumers. Using a similar updating rule of consumers' beliefs about product quality as in Li and Hitt (2010) and Jiang and Guo (2015), we assume that the next-period consumers form their quality beliefs by weighing the true quality q and the currently observed mean rating.

$$q^e_{t+1} = \omega q + (1 - \omega)E(r_t) \quad [10]$$

where $\omega \in [0, 1]$ represents the relative weight between a fully corrected quality measure q and the biased consumer reviews (r_1). When $\omega = 1$, consumers are fully rational and can overcome both self-selection biases that they form their second-period quality expectations based on the true quality q, while when $\omega = 0$, consumers cannot learn any information regarding product quality from the first-period reviews and they use the mean rating as second-period quality expectation.

If consumers can fully overcome both self-selection biases, $E(r_t) = q$. Equation [10] suggests that consumers' prior quality expectation should remain stable after the initial period, that is, $q^e = q$. By Proposition 1, we have the following Corollary:

Corollary 2:

(i) If consumers can fully overcome both self-selection biases, the rating series will be *stationary* with mean $q+\rho\sigma_q\lambda\left(\alpha(p,q)\right)+\sqrt{1-\rho^2}\sigma_q\Lambda_0-p$ and variance $\rho^2\sigma_q^2\left(1-\lambda\left(\alpha(p,q)\right)\left(\lambda\left(\alpha(p,q)\right)-\alpha(p,q)\right)\right)+\left(1-\rho^2\right)$ $\sigma_q^2\left(1+\Lambda_1-\Lambda_0^2\right)$.

(ii) If consumers can fully overcome both self-selection biases, the ratings are independent from each other.

In such a case, the expected difference of observed ratings at time t and at time t-1 should be a random variable whose value should be orthogonal with respect to the information set, which was known to consumers at time t-1 when they formed their expectation (in other words, the *ex post* rating error cannot be explained by past online product review information, such as the mean rating). If they are correlated, this indicates that consumers are not fully rational because they fail to use all information available to them at time t-1. If boundedly rational consumers form their prior quality expectations based on the mean rating without correcting for the self-selection biases, then the expected first difference of ratings is related to the mean of the observed ratings at t-1.

The above theoretical test framework is very similar to the rational expectation framework introduced by Muth (1961) and used by various studies (e.g., Forsells and Kenny 2002; Pesaran 1989). Fundamentally, this means that expectations are unbiased, and rational agents do not commit systematic and persistent errors when forming their quality expectation. In this paper, we test the orthogonality of rationality, meaning consumers revise their product quality expectation to reflect the flow of new information, and forecast errors should be uncorrelated; otherwise, consumers can improve their expectation by better using their past information.

Next, we first empirically tested the stationarity of the rating series r_i with real review data (Corollary 2) collected from Amazon.com (Appendix A). To ensure that products have a large enough number of reviews, we chose books with at least 100 reviews. We removed the first 20 reviews as the unsettling period. We conducted ADF test on the remaining review series for *each* product. The results are summarized in Table 3.

At the confidence level $p = .05$, the majority (75.3%) of products had a nonstationary rating series, implying that most consumers (or a single consumer at most times) are not able to fully capture unbiased information from online product reviews. While other reasons may cause the series of online product reviews to be nonstationary, we contend that bounded rationality can reasonably explain this nonstationary series, according to Corollary 2(i).

The above subsection showed that consumers cannot fully overcome the proposed self-selection biases in online product reviews. We now empirically further verify whether consumers are boundedly rational and form their prior quality expectations purely based on prior online product reviews, that is, $q_t^e = E(r_{t-1})$.

By Proposition 1, the expected difference of observed ratings at time t and at $t-1$, ($Rating_t - Rating_{t-1}$) contains biases that are correlated with prior quality expectations at $t-1$, that is, q_{t-1}^e. If boundedly rational consumers form their prior quality expectations based on the mean rating without correcting for the self-selection biases, $q_t^e = E(r_{t-1})$, then $E(Rating_t - Rating_{t-1})$ is related to the mean of the observed ratings at $t-1$ ($AvgRating_{t-1}$). Otherwise, if consumers fully correct for the self-selection biases, then $E(Rating_t - Rating_{t-1})$ will contain no biases, and will be independent of $AvgRating_{t-1}$.

We empirically verified whether consumers are boundedly rational (Proposition 1) by testing whether $Rating_t - Rating_{t-1}$ is correlated $AvgRating_{t-1}$ with a fixed-effect model presented in Equation [11] controlling for the natural log of the product's sales rank at Amazon at $t-1$, $\ln(SalesRank_{t-1})$. The natural log of the text length of the review $\ln(ReviewLength_{t-1})$ is used to control for information quality (Chevalier and Mayzlin 2006). We also controlled for $Price_{t-1}$, variance of product ratings $VarRating_{t-1}$, and dummy variables $Order_i$, which represents the position of each rating in the rating sequence (Godes and Silva 2012).

$$\begin{aligned} Rating_t - Rating_{t-1} &= \alpha_0 + \alpha_1 AvgRating_{t-1} + \\ \alpha_2\ln(SalesRank_{t-1}) &+ \alpha_3\ln(ReviewLength_{t-1}) + \\ \alpha_4 Price_{t-1} &+ \alpha_5 VarRating_{t-1} + \sum_i Order_i + \varepsilon \end{aligned} \qquad [11]$$

The model presented in Equation [11] is tested with book, DVD, and video data from Amazon separately, as shown in Table 4. The coefficients of $AvgRating_{t-1}$ are significant (marginally significant for Videos) and negative for all three categories, implying that consumers do not fully correct the self-selection biases in online product reviews when forming their prior quality expectations, which is the contraposition of Corollary 2(ii).

Integrating this finding with the nonstationarity test result, we conclude that *consumers are not completely rational* since they cannot fully overcome the self-selection biases in online product reviews.

Table 4 also shows that the change of ratings moves in the opposite direction with the mean rating of the last period. An

Table 3. Augmented Dickey-Fuller Test Results

	# of Items	Percentage of Items
Non-stationary	773	75.3%
Stationary	253	24.7%

Table 4. Consumer Awareness of Self-Selection Biases in Online Product Reviews

	Books	DVDs	Videos
$AvgRating_{t-1}$ (x10^3)	-17.4*	-45.1***	-8.4$^+$
ln($SalesRank_{t-1}$) (x10^3)	-0.4	-1.4	-0.5
ln($ReviewLength_{t-1}$) (x10^3)	-39.6***	-30.9***	-23.7***
$Price_{t-1}$ (x10^3)	20.7	5.8	2.2
$VarRating_{t-1}$ (x10^3)	-16.5**	-36.0***	-7.3$^+$
Adjusted R^2	6.8%	2.0%	2.2%

Note: Estimates of the coefficients of the fixed-effect dummies and the constant are omitted for brevity.
***p < 0.001, **p < 0.01, *p < 0.05, and ^+p < .10

extremely positive rating has a large deviation from the mean and inflates the mean rating; this results in a large correction that pulls down the mean in the next period. In addition, the variance of online product reviews also significantly corresponds to the change in ratings, implying a diversity effect. However, sales rank and price are not found to have a significant effect on the change in product ratings, implying that consumers do consider the price-induced acquisition bias in inferring product quality from online product reviews.

Combing these results with that of the sales forecasting model presented earlier, we conclude that consumers do *realize* the proposed self-selection biases and they *attempt to correct* for them by using other factors besides the mean rating to infer product quality. However, consumers *cannot perfectly correct for* these two self-selection biases due to bounded rationality. Also, since consumers are shown to infer product quality mainly from the mean rating without fully correcting for the self-selection biases, their surplus may suffer from the self-selection biases.

Firm's Pricing Decision in Response to Self-election Biases in Online Product Reviews

The firm's first-period pricing not only directly affects consumer demand and firm profit, but it also affects the consumers' ratings and their intention to write reviews. Those online product reviews will further shape the consumers' second-period quality expectations, which are often taken into account by the firm in making its second-period pricing decisions.

To study the firm's pricing strategy in response to the proposed self-selection biases in online product reviews, we started by choosing the second-period price to maximize the firm's second-period profit, before deciding on the first-period price to maximize the firm's total profit.

$$\max_{p_2} E(\pi_2) = p_2 * E\left(D\left(p_2, q_2^e\right)\right) = p_2 * E\left(\int_{\alpha\left(p_2, q_2^e\right)}^{+\infty} dF_\theta(\theta)\right) = p_2 * \left(1 - \Phi\left(\frac{p_2 - q_2^e - \mu_\theta}{\sigma_\theta + \rho\sigma_q}\right)\right) \quad [12]$$

The optimal second-period price p_2^* is obtained by the first-order-condition (after checking the second-order-condition) is

$$FOC_2(p_2) = \frac{\partial E(\pi_2)}{\partial p_2} = 1 - \Phi\left(\frac{p_2 - q_2^e - \mu_\theta}{\sigma_\theta + \rho\sigma_q}\right) - \frac{p_2}{\sigma_\theta + \rho\sigma_q}\phi'\left(\frac{p_2 - q_2^e - \mu_\theta}{\sigma_\theta + \rho\sigma_q}\right) = 0 \quad [13]$$

Holding the other parameters constant, we find that the optimal price has a positive relationship with q_2^e:

$$\frac{\partial p_2^*}{\partial q_2^e} = -\frac{\frac{\partial FOC_2}{\partial q_2^e}}{\frac{\partial FOC_2}{\partial p_2^*}} = -\frac{\left(\phi\left(\frac{p_2 - q_2^e - \mu_\theta}{\sigma_\theta + \rho\sigma_q}\right) + \frac{p_2}{\sigma_\theta + \rho\sigma_q}\phi'\left(\frac{p_2 - q_2^e - \mu_\theta}{\sigma_\theta + \sigma_q}\right)\right)}{\frac{\partial^2 FOC}{\partial p_2^{*2}}} > 0 \quad [14]$$

Accordingly, [14] suggests that, *ceteris paribus*, the firm's profit increases with the consumers' second-period product quality expectation. Having solved the optimal second period price, we traced back to the first period to consider the firm's optimal first-period price. For simplicity, we assumed no discounting for the second-period profit. The firm should choose an optimal first-period price p_1^* to maximize its overall profit for the two periods:

$$\begin{aligned}\max_{p_1} E(\pi) &= p_1 * E\left(D\left(p_1, q_1^e\right)\right) + p_2 * E\left(D\left(p_2, q_2^e\right)\right)\\ &= p_1 * E\left(\int_{\alpha\left(p_1, q_1^e\right)}^{+\infty} dF_\theta(\theta)\right) + p_2 * E\left(\int_{\alpha\left(p_2, q_2^e\right)}^{+\infty} dF_\theta(\theta)\right) \quad [15]\\ &= p_1 * \left(1 - \Phi\left(\tfrac{p_1 - q_1^e - \mu_\theta}{\sigma_\theta + \rho\sigma_q}\right)\right) + p_2 * \left(1 - \Phi\left(\tfrac{p_2 - q_2^e - \mu_\theta}{\sigma_\theta + \rho\sigma_q}\right)\right)\end{aligned}$$

The optimal first-period price can be derived by [15]:

$$\begin{aligned}FOC_1(p_1) = \tfrac{\partial E(\pi)}{\partial p_1} &= \left(1 - \Phi\left(\tfrac{p_1 - q_1^e - \mu_\theta}{\sigma_\theta + \rho\sigma_q}\right)\right)\\ &- \tfrac{p_1}{\sigma_\theta + \rho\sigma_q}\phi\left(\tfrac{p_1 - q_1^e - \mu_\theta}{\sigma_\theta + \rho\sigma_q}\right) + \tfrac{p_2}{\sigma_\theta + \rho\sigma_q}\phi\left(\tfrac{p_2 - q_2^e - \mu_\theta}{\sigma_\theta + \rho\sigma_q}\right)\tfrac{\partial q_2^e}{\partial p_1} = 0 \quad [16]\end{aligned}$$

The optimal price schedule (p_1^*, p_2^*) should satisfy both [13] and [16]. To solve for prices, we also need the connection between the first-period product ratings and second-period product quality expectation, that is, Equation [10]. We deduced from the examination of consumer rationality in the previous section that consumers draw their product quality expectations based on both true product quality and the mean of reviews left from previous consumers (r_1). The results in the previous section suggest that consumers do not achieve either of the above, thus $0 < \omega < 1$.

The firm's optimal price schedule (p_1^*, p_2^*) can be solved through the three simultaneous equations [10], [13] and [16]:

$$\begin{cases} 1 - \Phi\left(\tfrac{p_2 - q_2^e - \mu_\theta}{\sigma_\theta + \rho\sigma_q}\right) - \tfrac{p_2}{\sigma_\theta + \rho\sigma_q}\phi\left(\tfrac{p_2 - q_2^e - \mu_\theta}{\sigma_\theta + \rho\sigma_q}\right) = 0 \\ \left(1 - \Phi\left(\tfrac{p_1 - q_1^e - \mu_\theta}{\sigma_\theta + \rho\sigma_q}\right)\right) - \tfrac{p_1}{\sigma_\theta + \rho\sigma_q}\phi\left(\tfrac{p_1 - q_1^e - \mu_\theta}{\sigma_\theta + \rho\sigma_q}\right) \\ \quad + \tfrac{p_2}{\sigma_\theta + \rho\sigma_q}\phi\left(\tfrac{p_2 - q_2^e - \mu_\theta}{\sigma_\theta + \rho\sigma_q}\right)\tfrac{\partial q_2^e}{\partial p_1} = 0 \qquad [17] \\ q_2^e = \omega q + (1 - \omega)E(r_1) = \\ q + (1 - \omega)\left(\sqrt{1 - \rho^2}\,\sigma_q \Lambda_0\left(q_1^e\right) + \rho\sigma_q \lambda\left(\alpha\left(p_1, q_1^e\right)\right) - p_2\right) \end{cases}$$

When consumers can perfectly discern true product value from the ratings, that is, when their second-period quality expectation $q_2^e = q$, Equation [13] suggests that the optimal second-period price, product demand, and the firm's profit will be independent of consumers' first-period product quality expectation q_1^e. Otherwise, when the second-period consumers cannot fully correct the self-selection biases, their product quality expectation contains both self-selection biases in their prior ratings. Accordingly, the self-selection biases, together with the product quality expectation of first-period consumers will shape the firm's second-period pricing and profit.

Given that the solutions have very messy functional forms, we numerically solved [17] and plotted the numerical solutions to show the properties of the optimal prices, and the resulting product demand, firm profits, and consumer surplus in each period (Figure 5). For example, we set parameter values as $\mu_\theta = 0.5$, $\sigma_\theta = 0.2$, $q = 2$, $\sigma_q = 1$, $\rho = 0.5$, $\underline{\delta} = -2$, $\overline{\delta} = 0.7$, $\omega = 0.5$. We also plotted the numerical solutions for the scenario that consumers can form unbiased expectations for second-period quality, that is, $q_2^e = q$ (curves labeled with "_nb" in Figure 5). We also simulated the purchasing and reviewing decisions, as well as the updating of the product quality expectations of 1,272 consumers with specified distribution in both periods given the firm's prices. We summed up the net utility of all consumers to approximate consumer surplus, which is also shown in Figure 5.

We observe that the first-period price p_1, product demand d_1, and profit π_1 all increase with initial product quality expectation q_1^e, while the second-period quality expectation, price p_2, demand d_2, and profit π_2 all move in the opposite direction with q_1^e, except for the case when consumers form unbiased product quality expectations in the second period. When the initial product quality expectation q_1^e is low, the firm can charge a lower price to attract more consumers, which will result in more product reviews and potentially more positive reviews since their product quality expectation is low. As a result, a higher mean rating will increase the quality expectation of the second-period consumers, and the firm can charge a higher price to offset the lower profit margin in the first period, and to make more profits in the second period. To summarize, a lower first-period price might either directly improve future consumer ratings (direct impact) or indirectly boost consumers' online product reviews through stimulating the demand (indirect impact) by bringing another segment of consumers. The firm will choose the optimal price by balancing the impacts through consumer reviews, and the impact through profit margin. As q_1^e increases, the firm takes advantage of the high product quality expectations in the first period by charging a higher first period price.

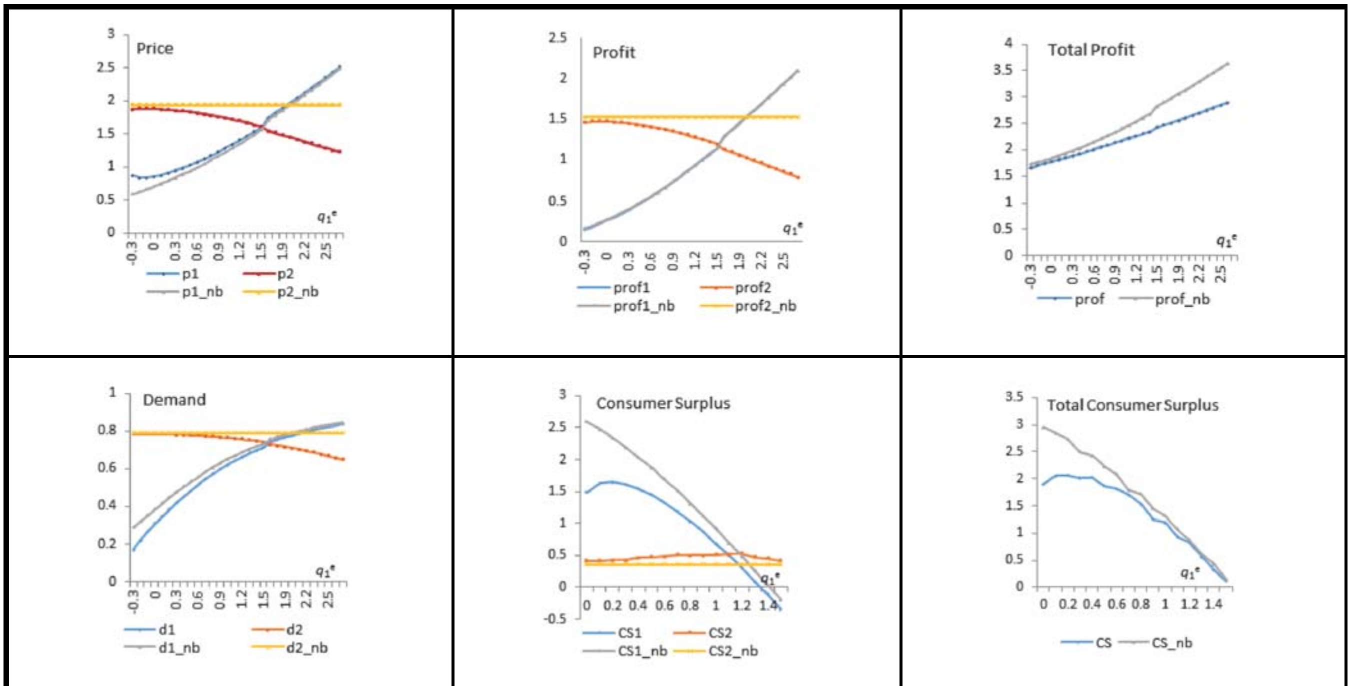


Figure 5. Changes in Optimal Price, Profit, and Demand of the Firm over Initial Quality Expectations (Curves for the scenario that consumers can form unbiased belief for second period quality are labled with "xx_nb")

Although the firm adjusts prices in response to the self-selection biases in online product reviews, our results imply that the self-selection biases may have a negative effect on the firm's profit and total consumer welfare. Product demand and profits of both periods are lower if consumers cannot rule out the self-selection biases in online product reviews when forming their product quality expectation. The second-period consumer surplus is higher when consumers do not fully correct the biases. However, the increase in second-period consumer surplus cannot offset the loss in the first-period consumer surplus. Accordingly, the total consumer surplus is reduced by the two proposed self-selection biases.

To demonstrate the role of the underreporting bias on the firm's profits and consumer surplus, we also plotted another set of graphs (Figure 6). The numerical results show the changes of the optimal pricing and profits with respect to the lower bound of consumers' reporting reviews $\underline{\delta}$ when $\overline{\delta}$ is fixed at 0.7. Figure 6 also shows that with the increase of $\underline{\delta}$ from –2 to 0, because more negative product reviews are reported, the firm will strategically reduce its prices in both periods, especially second-period price p_2. Also, the second period profit would decrease with an increase in $\underline{\delta}$. The first period profit is much smaller than the second-period profit, however, it increases slightly relative to the first-period profit as $\underline{\delta}$ goes up.

In sum, firms can strategically respond to the self-selection biases by adjusting their prices. For example, when consumers have low prior product quality expectations, the firm can lower the first-period price to attract more consumers to purchase and subsequently write more positive reviews, which will raise their second-period product quality expectation. Then the firm can charge a higher price in the second period to extract more profits. Nonetheless, since consumers cannot fully correct for the proposed self-selection biases, product demand and the firm's profits in both periods may suffer from the two self-selection biases. Total consumer surplus may also incur a loss due to the self-selection biases. We discuss the implications of these numerical results below.

Discussion

Key Findings and Contributions

We analytically modeled the consumer's two-stage (acquisition and reporting) decision-making process by examining the sources and nature of two self-selection biases (acquisition and underreporting). The dynamic process includes reading reviews to form prior product quality expectations, updating quality beliefs based on observable attributes, observing avail-

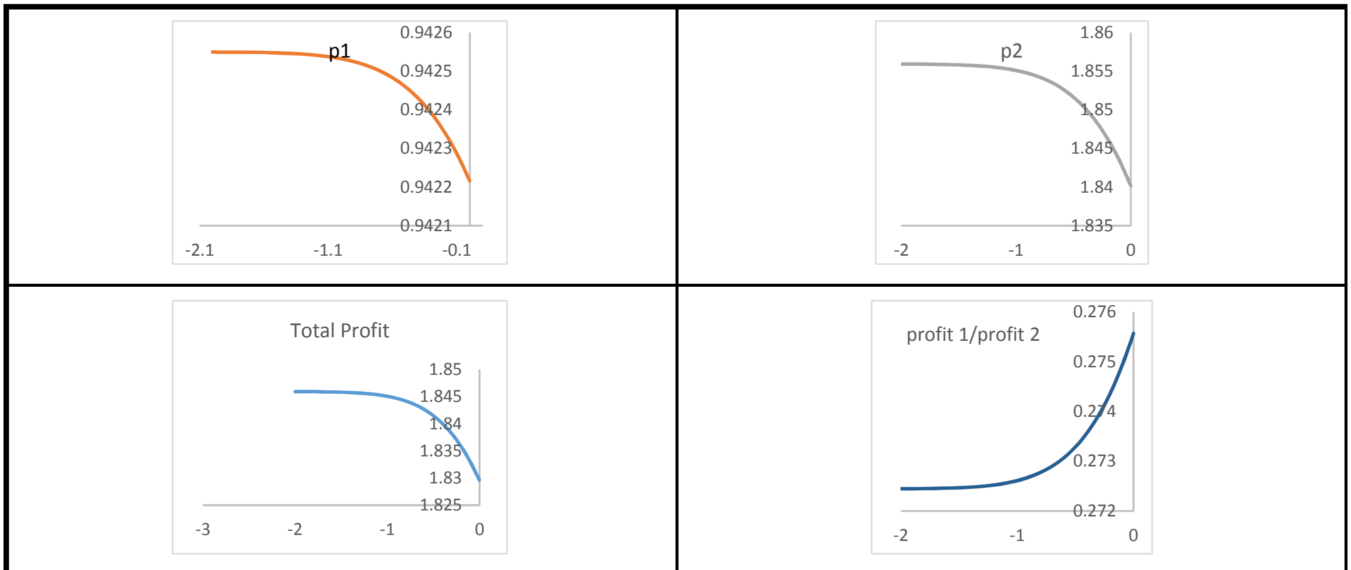


Figure 6. Changes of the Optimal Price, Profit, and Demand of the Firm over $\underline{\delta}$

able product reviews, making purchasing decisions, choosing to write a review based on actual satisfaction or dissatisfaction, and, when the new reviews are posted, how they influence the product quality expectations of subsequent consumers. Our analytical model and numerical results show that the mean rating may be a biased estimator of product quality that should not be used directly to either infer absolute product quality or relative quality across products.

Our results, in fact, show that the mean rating may not reflect the rank order of competing products. For example, for two products, one high quality (*mean* q = 2.2) versus one low quality product (*mean* q = 2.0). Assume w = 0.85 and X_H = 0.7 are fixed. For the high quality product, assume that underreporting bias is more serious (e.g., $X_L \geq 0.2$). In such a case, the mean rating will decrease as X_L increases. The maximum mean rating for the high quality product would be 1.2826 only. However, for the low quality product, assuming that underreporting bias is less serious. As long as its X_L is smaller than –1.6, the mean rating will always be larger than 1.2871. This represents a situation where the mean rating of a low quality product will have a higher mean rating than that of a higher quality product due to the less severe underreporting bias.

We analytically discuss the relative effect of the two self-selection biases on shaping consumer product quality expectations and affecting purchasing and reviewing decisions, and we show that relying on the mean rating without accounting for the self-selection biases decreases consumer surplus. Moreover, our consumer rationality models and empirical results suggest that consumers do realize the existence of the two self-selection biases when trying to infer quality based on online product reviews. While consumers attempt to account for the self-selection biases by using other distributional parameters of online product reviews besides the mean rating, they cannot perfectly correct for the two proposed self-selection biases due to bounded rationality. And we find that a firm's pricing decision is closely influenced by consumers' product quality expectations in the first period, as well as how likely consumers are to moan versus brag. We also found that product demand and the firm's profits in both periods suffer from the self-selection biases in online product reviews since consumers cannot fully correct for the self-selection biases. Our findings contribute to the literature by validating the broad consequences of self-selection biases in online product reviews. We show that self-selection biases are not completely mitigated by consumers due to bounded rationality, and they reduce their surplus by preventing them from inferring true product quality and making good purchasing decisions. In addition, product demand and the firm's profits may suffer from the two self-selection biases. Taken together, our key findings contribute to the emerging literature on online product reviews by demonstrating the negative effects of the self-selection biases on a broad spectrum of outcomes (product demand, consumer surplus, firm profits).

Implications for Theory

First, while there is an emerging interest in the nature and consequences of online product reviews (e.g., Dellarocas and

Wood 2008; Forman et al. 2008; Li and Hitt 2008), there is still no consensus as to whether, how, when, and why online product reviews can infer true product quality to predict product sales. Most studies use the mean rating as a predictor of product sales, assuming that it is an unbiased estimator of product quality. This assumption is challenged by our empirical results that validate the existence of an asymmetric, bimodal, positively-skewed (J-shaped) distribution of online product reviews. While the literature has speculated about the J-shaped nature of online product reviews (e.g., Admati and Pfleiderer 2004; Gao et al. 2015; Hu et al. 2009), this is the first study to analytically explain the J-shaped distribution of online product reviews due to the existence of the proposed self-selection biases (acquisition and underreporting). A key assumption in the literature is that, *ceteris paribus*, consumers will choose a product with a higher mean rating since either the mean rating reflects absolute product quality, or the relative difference in the mean ratings of two similar products would reflect their relative quality difference (e.g., Forman et al. 2008). Yet, this is not necessarily true and we show that the relationship between the mean rating and product quality is non-monotonic. Our study may thus partially explain the conflicting results in the literature that has often used the mean rating to predict sales, implying that the literature should go beyond single-point estimators (i.e., mean rating) to infer product quality. Our analytical results show that the mean rating contains both an unbiased product quality component and two self-selection biases. Theoretically, the unbiased quality component (mean rating) is only expected to explain future product sales. Accordingly, depending on the relative size of the unbiased and biased terms, the mean rating would vary in its ability to predict sales. This explains the mixed empirical findings in the literature regarding the predictive power of the mean rating on sales, predicated on the consumers' understanding and attempting to correct for the self-selection biases in online product reviews.

Banerjee and Fudenberg (2004) argue that self-selection biases do not inhibit social learning since "smart" consumers recognize and compensate for such biases by seeking additional information when estimating product quality. However, we show that consumers cannot fully correct for the self-selection biases because they are boundedly rational. Therefore, they cannot completely disregard the biases in online product reviews when inferring product quality. Similarly, Li and Hitt (2008) showed that consumers only *partially* account for self-selection biases. Hence, a key implication of our research is that consumers are *not* perfectly rational, and despite recognizing the existence of the two self-selection biases, they cannot fully overcome these two biases.

Our paper extends the literature on self-selection biases in online product reviews. Li and Hitt (2008) focused on acquisition bias and explained the dynamics of online product reviews due to changing consumer tastes as the product life cycle evolves. They showed that for each product, its early reviews tend to be more positive since early consumers are more positively predisposed toward the product. This may imply that the distribution of online product reviews may overcome this positive predisposition of the early consumers and reflect true product quality over time. However, the observed J-shaped distribution is not confined to early consumers, as our study empirically attests (Appendix A). This implies that the proposed self-selection biases in online product reviews do persist over time. Assuming a static distribution of consumer tastes over time, our paper posits that the review dynamics, which oscillate around the cumulative moving average, are driven by both proposed types of self-selection biases. Our paper also extends their finding regarding the consumers' understanding of the biases in the reviews. While they also recognize the consumers' limited ability to infer true product quality from biased online product reviews, we show that boundedly rational consumers are only able to realize their existence, but they are not able to fully overcome them to infer true product quality.

Finally, our paper contributes to the literature on the dynamics of online product reviews (e.g., Godes and Silva 2012; Moe and Schweidel 2012; Moe and Trusov 2011; Wang et al. 2011). These studies have examined the time series of online product reviews with a focus on social bias (or "diagnosticity assessment"); that is, whether the similarity in terms of the social attributes of online product reviews leads to an assimilation of product ratings over time. In our study, we focus on self-reporting bias and assume consumers truthfully report their quality assessments without influence from other reviewers. We build upon utility theory in the economics and satisfaction literatures (e.g., Anderson and Sullivan 1993) to assume that each consumer is a utility-maximizing agent who infers posterior quality information from online product reviews written by prior consumers, but she reports her own independent review based on her own individual experience.[10] We propose that consumer self-selection in acquisition and reporting, which cannot be completely corrected, may cause the dynamic effects on the rating series in online product reviews. In addition, rather than a strictly decreasing trend of the series of product ratings, based on our analytical results,

[10]This differs from papers on social biases (e.g., Moe and Trusov 2011; Wang et al. 2011) that are based on social influence theory (e.g., consumers' decisions about whether to review and what to report are influenced by other reviewers).

we theorize and empirically show that the move of a product rating is negatively correlated with the mean rating over time, revealing fluctuations around the cumulative moving average. In contrast to most of the literature that found the mean rating to reduce over time (Godes and Silva 2012; Moe and Schweidel 2012), we find that the mean rating may oscillate initially and may either go up or down over time. This pattern of ratings over time is supported with both numerical and real-life empirical examples. To our knowledge, this is the first study to show that acquisition bias and underreporting bias are both drivers of this trend of online product reviews that we observe in practice and, accordingly, our study brings new ideas and findings to the emerging literature on the dynamics of online product reviews. Our results indicate that the firm is likely to increase its prices as consumers' initial quality expectations increase. However, unless fully recovered, product demand and the firm's profit suffer from the proposed self-selection biases. Taken together, the implication of theory is that the two self-selection biases may have broad negative effects across the board, including demand for products, loss of consumer surplus, and reduction in the firm's profits.

Implications for Practice

This study has practical implications for (1) individual consumers, (2) online retailers, (3) companies that specialize in the collection and dissemination of online product reviews, and (4) product manufacturers.

First, consumers generally use the mean rating of online product reviews to infer product quality. This is intuitive since the mean is readily observed and easily understood. However, our results show that consumers must take into account the existence of self-selection biases, and they should adjust their estimation of product quality accordingly. First, consumers should take into consideration the standard deviation of the distribution. The larger the standard deviation of the distribution, the higher the uncertainty in inferring product quality and the greater the negative effect on quality. Second, consumers should also take into consideration the two modes of distribution of online product reviews. A higher upper mode or a smaller lower mode deflates product quality. Thus, product quality should be compensated more with a higher upper mode or a smaller lower mode. Third, price generally increases acquisition bias and inflates the observed mean rating. Therefore, for more expensive products, consumers should adjust the mean rating even lower to overcome the effect of acquisition bias. In summary, these findings suggest that consumers should compensate for acquisition and underreporting biases by using additional information beyond the mean rating of online product reviews to form their product quality expectations in order to improve their purchasing decisions and enhance their consumer surplus.

Second, since our model can help to estimate product quality and predict future sales, it can help the pricing decisions of online retailers who can shape product demand by adjusting consumer expectations according to online product reviews. For example, a product with a low mean rating will naturally have very low product demand due to the low quality expectation relative to its price, and its mean rating is unlikely to change over time as consumers are unlikely to purchase the product and write new product reviews. To raise product demand, retailers can strategically lower the price to attract new consumers. With a low prior quality expectation due to the low mean rating, new consumers will have a higher probability to be satisfied and write a more favorable review, which can boost the mean rating and increase future product sales.

Third, since many companies (e.g., Amazon.com, Barnes and Noble, Epinions.com, and BizRate.com) specialize on the collection, synthesis, and dissemination of online product reviews, having a more accurate method to inform consumers on product quality can be a differentiating factor. While they encourage their consumers to write and read product reviews, the simple mean rating they usually highlight is shown to be a biased estimator of product quality. Therefore, our results can be used by online review web sites to provide superior product information to their consumers. For example, complementing the reported mean rating with the distribution of online product reviews can help consumers better infer true product quality with such observable parameters regarding the distribution as the standard deviation and the modes. These parameters can not only help boundedly rational consumers avoid making poor purchasing decisions but also help alleviate the cognitive burden from "rational" consumers who strive to overcome the self-selection biases to infer product quality.

Finally, the results of this study can be used by product manufacturers to predict a the long-term success of a product based on reviews posted by early adopters. In doing so, product manufacturers can more reliably estimate which products to manufacture, and thus forge long-term agreements with suppliers and retailers. A better estimate of product quality and product demand can also be used by manufacturers and other players in a product's value chain to adjust production and distribution schedules, inventory management, and eventually the product's pricing to focus on products that are likely to have higher consumer demand in the future.

Limitations and Suggestions for Future Research

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

First, our results only use the numerical scores (number of stars) of online product reviews. However, consumers also write textual product reviews, and Archak et al. (2011) showed that the textual (verbal) characteristics of online product reviews affect sales beyond star ratings. Since consumers also read text reviews (e.g., Godes and Mayzlin 2004; Pavlou and Dimoka 2006), even if information overload prevents consumers from reading and processing the millions of words in text comments, future research could examine text comments. While the star ratings largely reflect the notion of text comments (Pavlou and Dimoka 2006), future research could examine biases in both the numerical and the textual aspects of online product reviews. The longitudinal study of biases in online product reviews could also be examined (e.g., Zheng et al. 2014).

Second, to reduce complexity, the analytical model assumed that consumers evaluate the product based on the reviews posted at the last period, and then make the purchasing and reviewing decisions at the same period. Under this assumption, we derive an unbiased estimator of product quality from online product reviews, while in reality there might be a delay in purchasing or reviewing. For future research, there are empirical ways to address such a question: (1) getting the archival data by collaborating with online vendors, such as Amazon, to find out the time elapse between a consumer making the purchase and when he/she writes the review, or (2) sending surveys to actual customers and asking for a direct answer. Future theoretical research could pursue a more natural setting that considers when consumers buy the product and when they post ratings at random time periods afterward.

Third, the five-point scale of star ratings may not be sensitive enough to fully capture the bimodal distribution. The DIP test (Appendix A) may not be able to distinguish between a bimodal distribution and a skewed unimodal distribution with only five points. Since the DIP test is a test of bimodality, it may not unequivocally conclude if there are two unimodal distributions or a bimodal distribution with only five points (similar to clustering five data points into one or two clusters). Thus, a scale with more anchors (such as 10 points) may be more appropriate.

Finally, while the field study (Appendix B) showed that online product reviews followed a roughly normal distribution, it is important to note that the simple field experiment was conducted among students, thus limiting the generalizability of our results. Future research could replicate our study with a randomized field experiment using a random sample of online consumers. Future research could also conduct a more controlled experiment with more products to compare "involuntary" reviews with Amazon's voluntary product reviews.

Concluding Remark

Online product reviews are a major informational source for consumers. Since online product reviews follow an asymmetric bimodal, positively-skewed (J-shaped) distribution due to the existence of two self-selection biases (acquisition and underreporting), the mean rating of online product reviews may be a biased estimator of product quality that reduces consumer surplus and forces consumers to include other distributional parameters to infer true product quality. While consumers do realize the two proposed self-selection biases and attempt to correct for them, they still cannot perfectly correct for these self-selection biases due to bounded rationality. Moreover, while firms adjust their prices in response to these self-selection biases, these biases reduce product demand and the firm's profits, while they also hurt consumer surplus. Taken together, these negative effects of the proposed self-selection biases call for future research to overcome self-selection biases in online product reviews.

Acknowledgments

We would like to thank the senior editor, Kartik Hosanagar, the associate editor, and the anonymous reviewers for their comments and suggestions that have significantly improved the quality of our manuscript.

References

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

Anderson, E. W. 1998. "Customer Satisfaction and Word of Mouth," *Journal of Service Research* (1:1), pp. 5-17.

Anderson, E. W., and Sullivan, M. 1993. "The Antecedents and Consequences of Customer Satisfaction for Firms," *Marketing Science* (12:2), pp. 125-143.

Archak, N., Ghose, A., and Ipeirotis, P. 2011. "Estimating the Socio-Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics," *Management Science* (57:8), pp. 1485-1509.

Arndt, J. 1967. "The Role of Product-Related Conversations in the Diffusion of a New Product," *Journal of Marketing Research* (4:3), pp. 291-295.

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

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

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

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

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

Chevalier, J., and Mayzlin, D. 2006. "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research* (43:3), pp. 345-354.

Clemons, E. K., Gao, G., and Hitt, L. 2006. "When Online Reviews Meet Hyperdifferentiation: A Study of the Craft Beer Industry," *Journal of Management Information Systems* (23:2), pp. 149-171.

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

Dellarocas, C., and Wood, C. 2008. "The Sound of Silence in Online Feedback: Estimating Trading Risks in the Presence of Reporting Bias," *Management Science* (54:3), pp. 460-476.

Dimoka, A., Hong, Y., and Pavlou, P. A. 2012. "On Product Uncertainty in Online Markets: Theory and Evidence," *MIS Quarterly* (36:2), pp. 395-426.

Duan, W., Gu, B., and Whinston, A. 2008. "The Dynamics of Online Word-of-Mouth and Product Sales—An Empirical Investigation of the Movie Industry," *Journal of Retailing* (84:2), pp. 233-242.

Forman, C., Ghose, A., and Wiesenfeld, B. 2008. "Examining the Relationship between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets," *Information Systems Research* (19:3), pp. 291-313.

Forsells, M., and Kenny, G. 2002. "The Rationality of Consumers' Inflation Expectations: Survey-Based Evidence for the Euro Area," European Central Bank Working Paper Series.

Gao, G., Greenwood, B.N. Agarwal, R., and McCullough, J. 2015. "Vocal Minority and Silent Majority: How Do Online Ratings Reflect Population Perceptions of Quality," *MIS Quarterly* (39:3), pp. 565-590.

Ghose, A., Smith, M. D., and Telang, R. 2006. "Internet Exchanges for Used Books: An Empirical Analysis of Product Cannibalization and Welfare Impact," *Information Systems Research* (17:1), pp. 3-19.

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

Godes, D., and Silva, J. 2012. "Sequential and Temporal Dynamics of Online Opinion," *Marketing Science* (31:3), pp. 448-473.

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

Hong, Y., and Pavlou, P. A. 2014. "Product Fit Uncertainty: Nature, Effects, and Antecedents," *Information Systems Research* (25:2), pp. 328-344.

Hu, N., Pavlou, P. A., and Zhang, J. 2006. "Can Online Reviews Reveal a Product's True Quality? Empirical Findings and Analytical Modeling of Online Word-of-Mouth Communication," in *Proceedings of the 7th ACM Electronic Commerce Conference*, Ann Arbor, MI.

Hu, N., Pavlou, P. A., and Zhang, J. 2009. "Why Do Product Reviews have a J-Shaped Distribution?" *Communications of the ACM* (52:10), pp. 144-147.

Jiang, Y., and Guo, H. 2015. "Design of Consumer Review Systems and Product Pricing," *Information Systems Research* (26:4), pp. 714-730.

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

Kuksov, D., and Xie, Y. 2010. "Pricing, Frills, and Customer Ratings," *Marketing Science* (29:5), pp. 925-943.

Li, X., and Hitt, L. M. 2008. "Self Selection and Information Role of Online Product Reviews," *Information Systems Research* (19:4), pp. 456-474.

Li, X., and Hitt, L. M. 2010. "Price Effects in Online Product Reviews: An Analytical Model and Empirical Analysis," *MIS Quarterly* (34:4), pp. 809-831.

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

Milgrom, P., and Roberts, J. 1986. "Price and Advertising Signals of Product Quality," *Journal of Political Economy* (94:4), pp. 796-821.

Moe, W., and Schweidel, D. A. 2012. "Online Product Opinions: Incidence, Evaluation, and Evolution," *Marketing Science* (31:3), pp. 372-386.

Moe, W., and Trusov, M. 2011. "The Value of Social Dynamics in Online Product Ratings Forums," *Journal of Marketing Research* (48:3), pp. 444-456.

Muth, J. F. 1961. "Rational Expectations and the Theory of Price Movements," *Econometrica* (29:3), pp. 315-335.

Pavlou, P. A., and Dimoka, A. 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), pp. 391-412.

Pesaran, M. H. 1989. "Consistency of Short-Term and Long-Term Expectations," *Journal of International Money and Finance* (8:4), pp. 511-516.

Thaler, R. 1980. "Toward a Positive Theory of Consumer Choice," *Journal of Economic Behavior and Organization* (1), pp. 39-60.

Wang, C., Zhang, X., and Hann, I-H. 2011. "Social Bias in Online Product Ratings: A Quasi-Experimental Analysis," Working Paper, Hong Kong University of Science and Technology.

Yadav, M., and Pavlou, P. A. 2014. "Marketing in Computer-Mediated Environments: Research Synthesis and New Directions," *Journal of Marketing* (78:1), pp. 20-40.

Zheng, Z., Pavlou, P. A., and Gu, B. 2014. "Latent Growth Modeling in Information Systems Research: Basic Tenets, Extension, and Practical Guidelines," *Information Systems Research* (25:3), pp. 547-568.

Zhu, F., and Zhang, X. 2010. "Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics," *Journal of Marketing* (74:2), pp. 133-148.

About the Authors

Nan Hu is an associate professor of Accounting at the School of Business at the Stevens Institute of Technology. He is also associated with the School of Management, Xi'an Jiaotong University, in Xi'an, Shan'xi, China. He received his Ph.D. from the University of Texas at Dallas. Nan's research focuses on investigating the value implications and market efficiency of both traditional and nontraditional information, using a combination of theories from accounting, finance, marketing, information economics, sociology, psychology, and computer science. Nan's research has appeared in *Journal of Accounting, Auditing, and Finance, Journal of Business Finance and Accounting, MIS Quarterly, Journal of Management Information Systems, IEEE Transactions on Engineering Management, Communications of the ACM, ACM Transactions on Management Information Systems, Decision Support Systems, Journal of Business Research,* and *Journal of Computer Security*, among others.

Paul A. Pavlou is the Milton F. Stauffer Professor of Information Technology and Strategy at the Fox School of Business at Temple University. He also serves as the Senior Associate Dean of Research, Doctoral Programs, and Strategic Initiatives, and directs the university-wide Big Data Institute at Temple University. Paul received his Ph.D. from the University of Southern California. He was ranked first in the world in publications in the two top MIS journals (*MIS Quarterly* and *Information Systems Research*) for 2010–2014. His work has been cited over 20,000 times by Google Scholar. Paul was recognized among the "World's Most Influential Scientific Minds" by Thomson Reuters based on analysis of highly cited authors during the 2002–2012 period. Paul won several best paper recognitions for his research, including the Maynard Award nomination for the "Most Significant Contribution to Marketing" in the *Journal of Marketing* in 2015, the *ISR Best Paper* award in 2007, and the 2006 *IS Publication of the Year* award.

Jie Jennifer Zhang is an associate professor of Information Systems in the College of Business, University of Texas at Arlington. She received her Ph.D. in Computer Information Systems from the Simon School of Business, University of Rochester. She employs analytical and empirical techniques to closely examine issues in advanced and business applications of information technologies, for example, trend and business impacts of social media, web advertising, consumer online behaviors, online channel selection and pricing strategies, software licensing policies, and business alliances. Her research appears in *Journal of Economics and Management Strategies, Information Systems Research, Journal of Management Information Systems,* and *Decision Support Systems*, among others. She has received grants and awards recognizing the impact of her research.

MIS Quarterly

RESEARCH ARTICLE

ON SELF-SELECTION BIASES IN ONLINE PRODUCT REVIEWS

Nan Hu
School of Business, Stevens Institute of Technology, 1 Castle Point Terrace,
Hoboken, NJ 07030 U.S.A. {nhu4@stevens.edu}

Paul A. Pavlou
Fox School of Business, Temple University, 1801 N. Broad Street,
Philadelphia, PA 19122 U.S.A. {pavlou@temple.edu}

Jie Zhang
College of Business, University of Texas, Arlington, 701 S. Nedderman Drive,
Arlington, TX 76019 U.S.A. {jiezhang@uta.edu}

Appendix A

The J-Shaped Distribution of Online Product Reviews

A random sample of product information and their corresponding consumer reviews were collected from Amazon in 2005 using Amazon Web Service (AWS) for more than 77,000 books, DVDs, and Videos from Amazon (Table A1).

Table A1. Descriptive Statistics for Amazon's Data

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

Figure A1 shows the distribution of the average rating for all books, DVDs, and videos on Amazon.com.

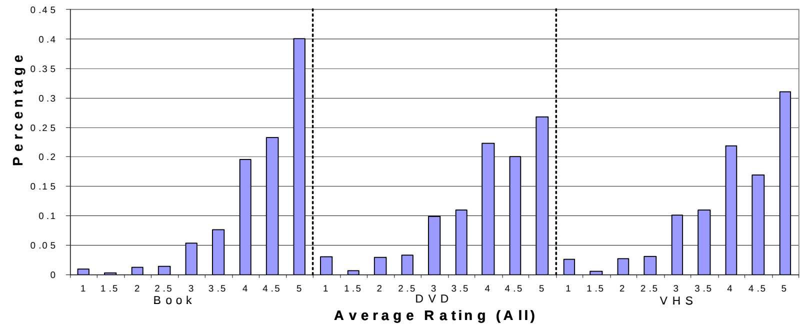


Figure A1. The Distribution of Online Product Reviews for All Books, DVDs, and Videos on Amazon.com

To verify that the J-shaped distribution does not vary over time, we split all Amazon's reviews into four equal groups (initial stage, early stage, late stage, final stage)[1] based on their posts. The J-shaped distribution persists (Figure A2).

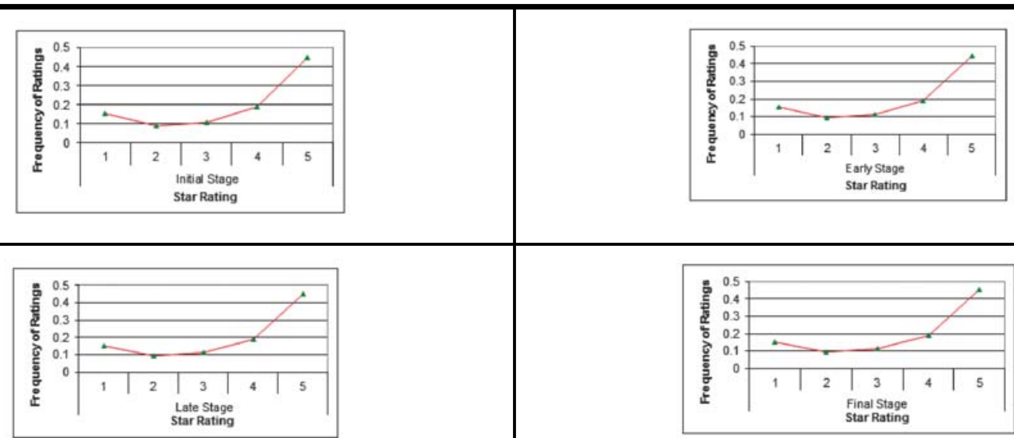


Figure A2. The Distribution of Online Product Reviews Over Time

Figure A3 shows the distribution of three randomly selected products in each of the three popular product categories with more than 2,000 reviews. The results show that these products also have a bimodal, asymmetric, left-skewed distribution, thus confirming that the observed J-shaped distribution is not due to the small number of product reviews.

[1]These four stages and their labels are proposed in a relative sense. Specifically, the initial stage reflects the earliest stage the product was first released; the final stage is the latest period. The J-shaped pattern still holds irrespective of periods and the absolute age of the reviews.

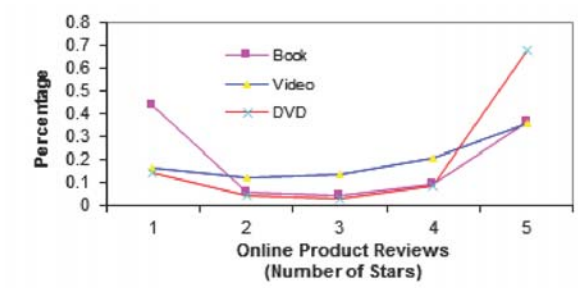


Figure A3. The Distribution of Online Product Reviews for Books, DVDs, and Videos with n > 2,000 Reviews

Figure A4 presents the distributions of online product reviews for freeware on Download.com with fewer than 20 reviews, which follow a bimodal, J-shaped distribution. This is a strong indication that the distribution is *not* normal. Also, to show that the J-shaped distribution applies to products with a different mean rating, Figure A4 also shows the distribution of products with a mean of 3.5-star (roughly in the middle) and 4-star (right hand side).[2]

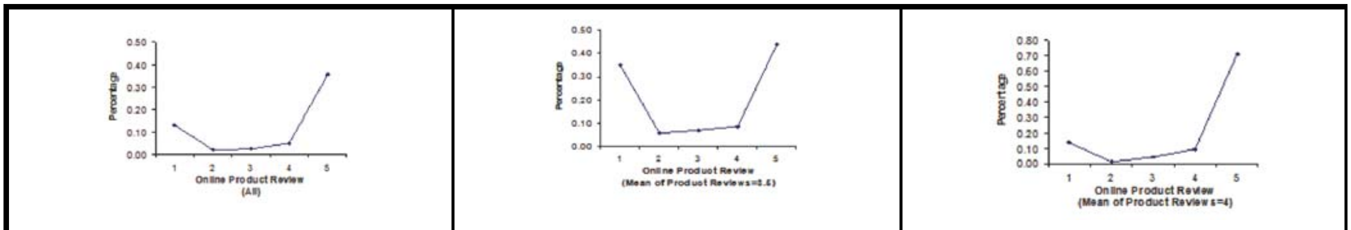


Figure A4. Distribution of Online Product Reviews of Freeware on Download.com (≤ 20 reviews)

Figure A5 shows the bimodal distribution of online reviews for products with a mean star rating of 3 and 4.[3]

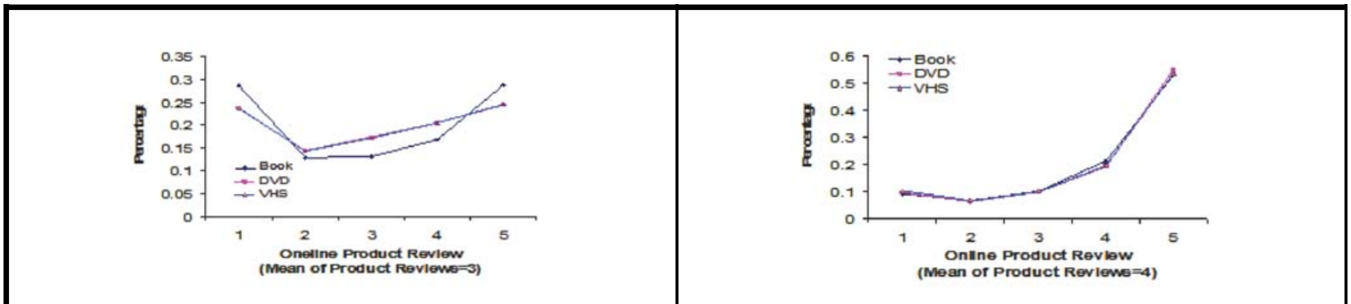


Figure A5. The Distribution of Books, DVDs, and Videos with Mean Rating of Three- or Four-Stars

[2]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 with a mean rating around 1 (e.g., between 0.9 and 1.1), into a group with mean = 1, and the results were very similar.

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

Figure A5 shows that as the mean rating increases, the U-shaped distribution becomes more left-skewed, thus turning into a J-shaped distribution. Besides the graphical inspections (Figures A1–A5), we formally tested whether the distribution of online product reviews is normal using the Kolmogorov-Snirnov test (Chakravarti et al. 1967), which examines if a sample comes from a normally distributed population.[4] This test at the individual product level showed that nearly all products do *not* follow a normal distribution.

To further test if the distribution of online product reviews for individual products is bimodal, the nonparametric DIP test (Hartigan and Hartigan 1985) was used. The DIP test is a measure of departure from unimodality; the DIP statistic for a unimodal distribution approaches zero, while the DIP statistic of a bimodal distribution approaches a positive constant.[5] We obtain the DIP statistics using ***R***.[6] The DIP test show that 90.17% of products have a distribution of online reviews that is neither unimodal nor normal. Virtually all products with a mean star rating between 1.5 and 4 stars do *not* follow a unimodal distribution. Even most products with a mean star rating around 5 stars do not follow a unimodal distribution.

To further establish the J-shaped distribution, we test the quadratic Equation A1 (Anderson 1998) using the following model:

$$f_{ij} = \alpha_{0j} + \alpha_{1j} s_{ij} + \alpha_{2j} s_{ij}{}^{2} + \sum \beta_{mj} x_{mj} + \varepsilon_{ij} \qquad \text{[A1]}$$

where f_{ij} is the number of product reviews with score i, for item j, $s_{ij} = i \in \{1, 2, 3, 4, 5\}$ is review score for product j, x_{mj} are other variables that might influence the rating of item j, such as price, mean rating and product category, and ε_{ij} is an error term. The null hypothesis to accept the bimodal distribution is given by H_0: $\alpha_1 < 0$ and $\alpha_2 > 0$.

To account for potential differences in product characteristics and means, we ran a fixed effect model by regressing the number of product reviews on the star rating (number of stars). As a robustness check, we ran separate regressions for different groups composed of products from the same product category and with similar mean rating of product reviews, and we then estimated the mean coefficient across these categories. The results are qualitatively the same. The results when all products are pulled together show a significant negative $\alpha_1 = -40.54$ and a significant positive value $\alpha_2 = 9.09$. Therefore, the estimated quadratic curve of [A1] is symmetric in terms of the rating $s_i = 2.2$, which lies to the left of the median point of 3, implying that online product reviews for virtually all products collected from Amazon within the range of 1–5 star ratings have a J-shaped (left skewed bimodal) distribution. As an additional robustness check, we ran this J-shaped test on individual product level. Our results indicate that 83.1% of books, 82.2% of DVD, and 76.0% of VHS follow a J-shaped distribution.

---

[4] We also employed the Cramer-von Mises (Thode 2002) and the Anderson-Darling (Stephens 1974) tests with similar results.

[5] Besides DIP test, other nonparametric tests of unimodality are available, such as 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 as the excess mass statistic is exactly twice the DIP statistic (Cheng and Hall 1998). However, the DIP test is simpler and more conservative (Cheng and Hall 1998; Henderson et al. 2000). Therefore, if the DIP test shows a large percentage of online product reviews to have a bimodal distribution, the other tests are likely to provide even more pronounced results.

[6] For more information on ***R***, see http://www.r-project.org.

Appendix B

Lab Experiment on Self-Selection Biases in Online Product Reviews

We asked 218 subjects to review four products (music CD, movie DVD, Access software, IS textbook) as well as their review, purchase intentions, and purchase importance on a 1–5 scale. These products were chosen to vary in terms of product category (music, movies, software, textbook), prior ownership, familiarity, importance, and price level ($10–$250). For each product, we assured that the subjects were familiar with each product. They were asked to hear all twelve 30-second clips of the music CD, and they were also asked to watch movie "Titanic" if they did not. Subjects used Access as part of a required class assignment, while for them the IS textbook was a required class textbook. Subjects were asked to rate the product and also report whether (1) they had already owned the product, (2) the importance of the product to them, and (3) their intention and passion to report a product review. The purpose is to compare the distribution of online product reviews from almost all respondents in the lab experiment with that of reviews on Amazon.com. Table B1 shows the number of respondents for each product, and descriptive statistics of their responses. The response rate of over 92% shows that the products were reviewed by almost all participants in our sample, and nonresponse bias tests showed that the nonrespondents did not differ from the respondents.

Table B1. Sample Characteristics

Product	Number of Subjects	Number of Reviews on Amazon	Prior Ownership (Percentage)	Intention to Review (Mean – STD)	Purchase Importance (Mean – STD)
Music CD	197	157	8%	2.03 (0.95)	2.23 (0.99)
Movie DVD	199	2107	35%	2.25 (1.00)	2.69 (1.16)
Software	203	10	66%	2.18 (0.97)	2.59 (1.06)
Textbook	201	13	83%	2.54 (1.03)	3.67 (1.17)

Amazon's and the experiment's mean ratings for each of the four products are quite different (Table B2). While the music CD and textbook are rated higher on Amazon ($p < .001$), the movie DVD is rated higher among the experiment's respondents; finally, the mean rating for the Access software is roughly the same between Amazon and the experiment.

Table B2. Differences in Mean Star Ratings of Product Reviews Survey Versus Amazon

Product	Field Study	Amazon	Equality Test (p-value)
Music CD	3.25	3.90	0.0000
Movie DVD	4.09	3.56	0.0000
Access Software	3.53	3.60	0.7931
IS Textbook	3.51	4.79	0.0000

Besides graphical differences (Figures 1 and 2), we specified a system of equations to isolate the self-selection biases:

$$Rating = \alpha_0 + \alpha_1 Ownership + \alpha_2 Intention + \alpha_3 Importance + \Sigma_{i=1}^{3} \alpha_i ProductDummy_i + \varepsilon \quad \text{[B1]}$$

$$Intention = \beta_0 + \beta_1 Rating + \beta_2 Rating^2 + \beta_3 Importance + \Sigma_{i=1}^{3} \beta_i ProductDummy_i + \eta \quad \text{[B2]}$$

where *Rating* = The respondent's star rating on a five-point Likert-type scale anchored between one star and five stars.
Ownership = Binary variable whether the resondent already owns the product.
Intention = The respondent's intention to write a product review at Amazon.com on a five-point scale.
Importance = The respondent's assessment of how salient the purchase is on a five-point scale.
ProductDummy = Represents the fixed effects due to potential differences across the three categories.

In Equation [B1], we used *Ownership* as a proxy for acquisition bias. Equation [B1] summarizes the predictors of the star rating. The utility theory suggests that prior ownership is expected to increase the mean rating. In fact, the mean rating of subjects who already owned the product was significantly higher than those who do not ($p < .05$) (Table B3). Purchase importance was controlled for its positive effect on the star rating since consumers who perceive the purchase to be important are more likely to be positively predisposed toward the product and to write a positive review.

In Equation [B1], intention to write a review was used as a proxy for underreporting bias. If consumers are more likely to write a review when they are either extremely satisfied or dissatisfied, *Intention* is positively correlated with *ExtremeStarRating* (a dummy variable =1 when consumers leave a one-star or five-star rating and 0 otherwise). Wald's test in Table B4 based on Equation [B3] supports this positive correlation ($\chi^2 = 40.98$, $p < .0001$). We estimate Equation [2] in which the subjects' intention to report a review on Amazon is determined by the subject's star rating. *Rating* and *Rating*² are included to account for a potential nonlinear effect of the respondent's star rating and her intentions to write a review.

Table B3. Differences in Star Ratings of Online Product Reviews Based on Prior Ownership

Product	Prior Ownership (mean)	No Ownership (mean)	Difference: Sign	Difference: Difference	t-value	p-value
Music CD	4.20	3.18	+	1.02	4.3600	<.0001
Movie DVD	4.26	4.00	+	0.26	2.2000	0.029
Access Software	3.64	3.30	+	0.34	2.8600	0.0047
IS Textbook	3.58	3.20	+	0.38	2.8700	0.0046

Table B4. Likelihood of Writing an Extreme Product Review

	Coefficient	Wald Chi-Square	p-value
Intercept	-3.68	105.24	<.0001
Intention	0.68	40.98	<.0001
DVD Dummy	1.31	20.74	<.0001
Software Dummy	-0.04	0.01	0.9081
Textbook Dummy	-1.13	7.90	0.0050

Since the *Intention* variable in Equation [B1] is a linear combination of other variables in Equation [B2], we adopted the limited-information maximum likelihood (LIML) estimation to simultaneously estimate the system of equations for acquisition bias and underreporting bias.[7] The results are reported in Table B5.[8]

$$ExtremeStarRating = \gamma_0 = \gamma_1 Intention + \sum_{i=2}^{4} \gamma_i ProductDummy_i + \varepsilon \qquad \text{[B3]}$$

[7]Asymptotically, 2SLS and LIML estimators have the same distribution (Anderson 2005). Even though it is easier to compute 2SLS, LIML was used because (1) the parameter estimation method of simultaneous equation models was based on ML that is commonly believed to yield superior estimators (Anderson 2005); (2) LIML takes into account the covariances of the error terms; (3) 2SLS estimator treats the components of β asymmetrically, which runs contrary to simultaneous equations (Anderson 2005, p. 9).

[8]As a robustness check, besides estimating the system of Equations 1 and 2, we also estimated these equations independently using both OLS and logistic regression. Furthermore, we estimated Equation 2 with the sample composed of all respondents, or those respondents who already owned the product before. All of these tests have qualitatively the same results.

Table B5. Acquisition Bias and Underreporting Bias Based on System Equations [1] and [2]			
Acquisition Bias		**Underreporting Bias**	
Variable	**Parameter *α***	**Variable**	**Parameter *β***
Intercept	2.23***	Intercept	1.10***
Prior Ownership	0.13**	Rating	-0.19*
Intention to Write Review	0.05	Rating²	0.04**
Purchase Importance	0.28***	Purchase Importance	0.44***
CD_Dummy	0.28***	CD_Dummy	0.13*
DVD_Dummy	0.94***	DVD_Dummy	0.07
Software_Dummy	0.37***	Software_Dummy	0.11
N	800	N	800
Adjusted R²	28.9%	Adjusted R²	29.9%
DW	1.98	DW	2.07

****p* < 0.001; ***p* < 0.05; **p* < 0.10

For acquisition bias, after controlling for purchase importance, prior ownership (α_1 = 0.13, $p < 0.05$) was positively linked to rating, and intention to report a review was not significant ($\alpha_3 = 0.05$, $p > 0.10$). For underreporting bias, the negative coefficient of *Rating* was marginally significant ($\beta_1 = -0.19$, $p < 0.10$), and the *Rating*² coefficient ($\beta_2 = 0.04$, $p < 0.05$) was positive and significant. That implies that consumers are more likely to write a review when they are either satisfied or dissatisfied, while they are the least likely to report a review when their star rating is moderate ($-\frac{\partial \beta_2}{2\partial \beta_1} = 2.4$ stars) according to Equation [B2]. These results reinforce that acquisition bias (reflected through prior ownership) and underreporting bias (reflected through higher intentions to write a product review) result in the observed J-shaped distribution.

Figure B1 shows that there is a stark contrast between the means of Amazon's online product reviews and those of the lab experiment reviews for the exact same four products (music, movie, software, textbook). Amazon's online product reviews resemble a J-shaped distribution, whereas the lab experimental data follow a unimodal, roughly normal distribution. Furthermore, while the majority of Amazon's online product reviews are extreme or polarized (one-star or five-star), the majority of the lab experiment's product reviews (over 90%) are moderate (two-star, three-star, or four-star). Finally, while Amazon's online product reviews are mostly positive (five-star), the lab experiment's results are balanced across all star ratings between one-star and five-stars.

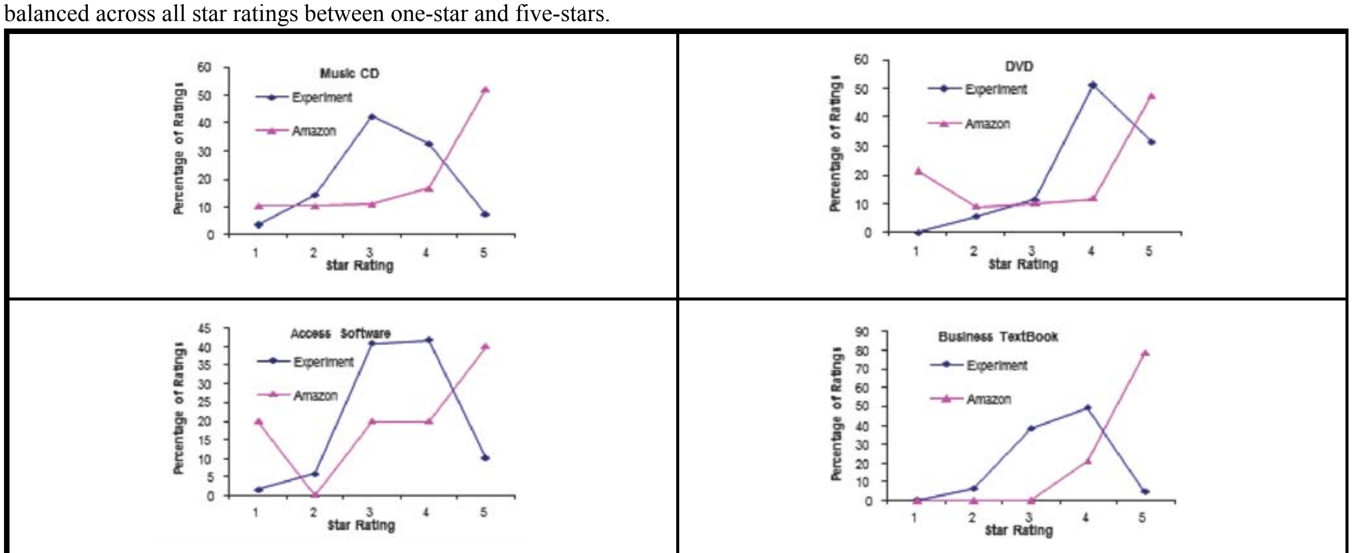


Figure B1. Comparison of Amazon's and the Lab Experiment's Distributions of Online Product Reviews

In addition to showing that the lab experiment's data follow a normal distribution (Figure B1), we isolated the two self-selection biases by plotting the distribution of the respondents with prior ownership (capturing acquisition bias) and respondents with high intentions (≥ 3) to write an online review on Amazon.com (capturing underreporting bias). As shown in Figure B2 for the movie DVDs (the other products follow a similar pattern and are omitted for brevity), prior ownership shifts the distribution toward higher ratings. Besides, selecting those respondents with high intentions to write a review largely omits the moderate star ratings, resulting in a distribution that resembles Amazon's observed J-shaped distribution. In sum, the combination of the two self-selection biases is shown to jointly shift a normal distribution of all respondents to a left-skewed bimodal distribution, which resembles Amazon's J-shaped distribution.

Figure B2. Distribution of Online Product Reviews from Amazon Versus Experimental Field Study Results

Appendix C

Comparison of Models with and Without Self-Selection Biases in Online Product Reviews

To compare the proposed "dual mode" model against the three competing models (mean average model, weighted mean average model, and extreme rating controlled model), we randomly selected another 10,000 books, DVDs, and videos from our Amazon sample starting in July of 2005, with the SAS random function. For each product, we collected its price, sales rank, and all online product reviews for several months on three-day intervals.[9] Therefore, since we have panel data from July 2005 to January 2006, we can compare which model has the highest power in terms of predicting future product sales using longitudinal secondary data. The models, whose predictive validity is shown in Table C1, are compared below.

Proposed Dual Mode Model

$$\ln(SalesRank_{i+1}) = \beta_0 + \beta_1 AvgRating_t + \beta_2 X_{Lt} + \beta_3 X_{Ut} + \beta_4 StdevRating_t + \beta_5 \ln(SalesRank_t) + \beta_6 \ln(Price_t) + \beta_7 \ln(NumRev_t) + \beta_8 Book\_Dummy + \beta_9 DVD\_Dummy + \varepsilon_t$$

Model 1 (Simple Mean)

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

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*)

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

Model 2c (Weighted Mean c*)

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

$*Weighted\_Mean\_Rating_{1i} = Avg((HelpfulReviews/TotalReviews) * ReviewRating)$
$*Weighted\_Mean\_Rating_{2i} = Sum((HelpfulReviews/TotalReviews) * ReviewRating)/Sum(HelpfulReviews/TotalReviews)$
$*Weighted\_Mean\_Rating_{3i} = Sum(HelpfulReviews * ReviewRating) / Sum(HelpfulReviews)$

[9] Ideally, we would like to collect data on a daily basis because price, sales, and online product reviews change on a daily basis. However, due to the instability of Amazon's web service, it often takes more than three days to collect a batch of data. Thus, to ensure that we get clean copy for each batch of data, we used a three-day instead of a one-day interval.

Model 3 (One-Star and Five-Star Model)

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

As shown in Table C1, controlling for previous sales rank,[10] price, and the total number of product reviews, the model with self-selection controlled explains a substantial amount of the variance (R^2 adjusted = 77.29%) in future product sales, which is significantly higher than all other models ($p < .0001$). The mean of the online product reviews has a significant effect, explaining .16% of the variance in future product sales. The two dual modes X_L and X_U *also* have significant effects ($p < .001$),[11] explaining .64% and .33% of the variance in future product sales, respectively. 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 the literature (e.g., Chevalier and Mayzlin 2006) that suggests that consumers pay more attention to negative reviews (which are generally captured by X_L) compared to positive reviews (which are captured by X_U). Finally, the STD is statistically significant (p<.05), explaining .034% of the variance. This is consistent with Clemons et al. (2004) who showed that the variance of online product reviews affects future sales.

Table C1. Model Comparisons

	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*	
X_L	0.0573***					
X_U	-0.0685***					
STD†	0.0086*					
Percent(1-star reviews)						0.1088*
Percent(5-star reviews)						-0.0452
Ln (Current Sales Rank)	0.7114***	0.7206***	0.7218***	0.721***	0.7215***	0.7208***
ln(Price)	0.0780***	0.0999***	0.100***	0.0989**	0.0986***	0.099***
ln (# of Product Reviews)	-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 R^2	**77.29%**	**75.18%**	**75.17%**	**75.25%**	**75.24%**	**75.17%**
Difference in R^2		**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

†STD is weighted by helpvote/totalvote.
****p* < .001; ***p* < .01; **p* < .05; +*p* < .10. All *p*-values are two-sided.

We used the following equation for calculating the significance between two regression models:

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

[10]The time difference between $t + 1$ and t is 130 days. We also tested other time lag values (e.g., 100 days, 110 days), which yielded similar results.

[11]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.

where K_x is the number of independent variables in the proposed Model X
K_i is the number of independent variables in the competing Model I
N is the sample size

There are several criteria that can be used to choose among competing models, such as the Adjusted R^2, 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 and Kilian (2001), no criterion is necessarily superior. For simplicity, we evaluated the performance of the various competing models by comparing their adjusted R^2, which is the most widely used criterion for model comparison. 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. For non-nested model comparison, we also used the Davidson-MacKinnon J test, which showed similar results.

The proposed self-selection controlled model explains at least 2% higher variance compared to the five competing models (which roughly explain about the same variance).[12] This difference in variance explained is statistically significant ($p < .0001$), as the F-tests in Table C1 attest. Besides the high F-values that denote 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,[13] the number of online product reviews explains 4.55%, the product dummies explain 7%, and price only explains .11%. Given these influential control variables, the variance explained by the new proposed independent variables (X_L, X_U, STD) is also substantial from a practical standpoint, attesting to the need for including these distributional parameters when predicting future product sales.

Table C2. Comparison between Proposed Modes and Percentage of Polarized Reviews

	Regression Model
Mean_Product Reviews	-0.0906*
X_L	0.0563***
X_U	-0.0688***
STD	0.0083*
% 1-star reviews	-0.1895$^{N/S}$ (p = .1739)
% 5-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 R²	77.29%
N	7573

[12]Interestingly, none of the three proposed weighted means of online product reviews is superior to the simple mean rating. Perhaps this is because consumers only observe the simple mean and do not otherwise process the number of useful reviews.

[13]We also ran the same regression models (Table C1) by omitting the current sales rank as a control variable, and the results were very similar (Model X outperformed all others by over 2%). However, since the variance explained in future sales is substantially lower (circa 35%) when omitting current sales rank, we only report the results with all control variables.

While the proposed prediction model with the proposed X_L and X_U modes is superior to the one with the polarized (one-star and five-star) reviews (Model 3), we still wanted to have a direct comparison of their joint impact. Therefore, we ran a regression model in which we included both X_L and X_U and also the percentage of one-star and five-star reviews (Table C2). 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 one-star and five-star reviews, allowing us to simultaneously include them in a regression model. As shown in Table C2, both the percentage of polarized (1-star and 5-stars) reviews 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.

Appendix D

Proofs

Proof of Lemma 1:

With a window as described in Equation [6], the density function of reviews is derived as

$$f(q_i|\theta_i)=\begin{cases} 0 & if\ \underline{\delta}\le u_i-E(u_i|\theta_i)\le\bar{\delta} \\ \dfrac{\phi\left(\frac{q_i-q^e-\rho\sigma_q(\theta_i-\mu_\theta)/\sigma_\theta}{\sqrt{1-\rho^2}\sigma_q}\right)\Big/\sqrt{1-\rho^2}\sigma_q}{\Phi\left(\frac{\underline{\delta}+(q^e-q)}{\sqrt{1-\rho^2}\sigma_q}\right)+\left(1-\Phi\left(\frac{\bar{\delta}+(q^e-q)}{\sqrt{1-\rho^2}\sigma_q}\right)\right)} & Otherwise \end{cases} \quad \text{[D1]}$$

The expected rating from consumer i who purchases the product with $E(u_i|\theta_i)\ge 0$ that is, $\theta_i\ge\alpha(p,q^e)$ with $\alpha(p,q^e)=\frac{\sigma_\theta(p-q^e)+\rho\sigma_q\mu_\theta}{\sigma_\theta+\rho\sigma_q}$, is derived by

$$E(r_i\,|\,\theta_i\ge\alpha)=E(q_i\,|\,\theta_i\ge\alpha)=\int q_i f(q_i\,|\,\theta_i)dq_i$$

$$=q+\rho\frac{\sigma_q}{\sigma_\theta}(\theta_i-\mu_\theta)+\left(\int_{-\infty}^{\underline{\delta}+q^e+\rho\frac{\sigma_q}{\sigma_\theta}(\theta_i-\mu_\theta)}\frac{\frac{q_i}{\sqrt{1-\rho^2}\sigma_q}\phi\left(\frac{q_i-q-\rho\frac{\sigma_q}{\sigma_\theta}(\theta_i-\mu_\theta)}{\sqrt{1-\rho^2}\sigma_q}\right)}{\Phi\left(\frac{\underline{\delta}+q^e-q}{\sqrt{1-\rho^2}\sigma_q}\right)+\left(1-\Phi\left(\frac{\bar{\delta}+q^e-q}{\sqrt{1-\rho^2}\sigma_q}\right)\right)}dq_i+\right.$$

$$\left.\int_{\bar{\delta}+q^e+\rho\frac{\sigma_q}{\sigma_\theta}(\theta_i-\mu_\theta)}^{+\infty}\frac{\frac{q_i}{\sqrt{1-\rho^2}\sigma_q}\phi\left(\frac{q_i-q-\rho\frac{\sigma_q}{\sigma_\theta}(\theta_i-\mu_\theta)}{\sqrt{1-\rho^2}\sigma_q}\right)}{\Phi\left(\frac{\underline{\delta}+q^e-q}{\sqrt{1-\rho^2}\sigma_q}\right)+\left(1-\Phi\left(\frac{\bar{\delta}+q^e-q}{\sqrt{1-\rho^2}\sigma_q}\right)\right)}dq_i\right)$$

$$=q+\rho\frac{\sigma_q}{\sigma_\theta}(\theta_i-\mu_\theta)+ \quad \text{[D2]}$$

$$\sqrt{1-\rho^2}\sigma_q\left(\int_{-\infty}^{\frac{\underline{\delta}+q^e-q}{\sqrt{1-\rho^2}\sigma_q}}\frac{\frac{q_i-q-\rho\frac{\sigma_q}{\sigma_\theta}(\theta_i-\mu_\theta)}{\sqrt{1-\rho^2}\sigma_q}\phi\left(\frac{q_i-q-\rho\frac{\sigma_q}{\sigma_\theta}(\theta_i-\mu_\theta)}{\sqrt{1-\rho^2}\sigma_q}\right)}{\Phi\left(\frac{\underline{\delta}+q^e-q}{\sqrt{1-\rho^2}\sigma_q}\right)+\left(1-\Phi\left(\frac{\bar{\delta}+q^e-q}{\sqrt{1-\rho^2}\sigma_q}\right)\right)}d\left(\frac{q_i-q-\rho\frac{\sigma_q}{\sigma_\theta}(\theta_i-\mu_\theta)}{\sqrt{1-\rho^2}\sigma_q}\right)+\right.$$

$$\left.\int_{\frac{\bar{\delta}+q^e-q}{\sqrt{1-\rho^2}\sigma_q}}^{+\infty}\frac{\frac{q_i-q-\rho\frac{\sigma_q}{\sigma_\theta}(\theta_i-\mu_\theta)}{\sqrt{1-\rho^2}\sigma_q}\phi\left(\frac{q_i-q-\rho\frac{\sigma_q}{\sigma_\theta}(\theta_i-\mu_\theta)}{\sqrt{1-\rho^2}\sigma_q}\right)}{\Phi\left(\frac{\underline{\delta}+q^e-q}{\sqrt{1-\rho^2}\sigma_q}\right)+\left(1-\Phi\left(\frac{\bar{\delta}+q^e-q}{\sqrt{1-\rho^2}\sigma_q}\right)\right)}d\left(\frac{q_i-q-\rho\frac{\sigma_q}{\sigma_\theta}(\theta_i-\mu_\theta)}{\sqrt{1-\rho^2}\sigma_q}\right)\right)$$

$$=q+\rho\tfrac{\sigma_q}{\sigma_\theta}(\theta_i-\mu_\theta)+\sqrt{1-\rho^2}\sigma_q\Lambda_0(q^e)-p \qquad where\ \Lambda_0(q^e)=\frac{\phi\left(\frac{\bar{\delta}+(q^e-q)}{\sqrt{1-\rho^2}\sigma_q}\right)-\phi\left(\frac{\underline{\delta}+(q^e-q)}{\sqrt{1-\rho^2}\sigma_q}\right)}{\Phi\left(\frac{\underline{\delta}+(q^e-q)}{\sqrt{1-\rho^2}\sigma_q}\right)+\left(1-\Phi\left(\frac{\bar{\delta}+(q^e-q)}{\sqrt{1-\rho^2}\sigma_q}\right)\right)}$$

Q.E.D.

Proof of Proposition 1:

Following [D2], we derive the expected review score from a consumer i with prior quality expectation q^e:

$$E(r_i) = E\left(E(q_i - p)\right)$$
$$= E\left(q + \rho\frac{\sigma_q}{\sigma_\theta}(\theta_i - \mu_\theta) + \sqrt{1-\rho^2}\sigma_q\Lambda_0\left(q^e\right) - p\right) \quad \text{[D3]}$$
$$= q + \rho\sigma_q\frac{\phi\left(\alpha\left(p,q^e\right)\right)}{1-\Phi\left(\alpha\left(p,q^e\right)\right)} + \sqrt{1-\rho^2}\sigma_q\Lambda_0\left(q^e\right) - p$$

If we let $\lambda(x) = \frac{\phi(x)}{1-\Phi(x)}$, then [D3] can be expressed as

$$E(r) = q + \sqrt{1-\rho^2}\sigma_q\Lambda_0\left(q^e\right) + \rho\sigma_q\lambda\left(\alpha\left(p,q^e\right)\right) - p \quad \text{[D4]}$$

$\rho\sigma_q\lambda\left(\alpha\left(p,q^e\right)\right)$ and $\sqrt{1-\rho^2}\sigma_q\Lambda_0\left(q^e\right)$ are the acquisition bias and the underreporting bias, respectively. If $\rho \neq 1$, getting rid of the underreporting bias requires $\Lambda_0 = 0$, which can only be achieved when $\underline{\delta}$ and $\overline{\delta}$ are symmetric in terms of the difference between the realized quality and prior, that is, $\overline{\delta} + \underline{\delta} = 2(q - q^e)$, if $\overline{\delta} = \underline{\delta} = 0$ is impossible due to the prevalence of underreporting bias.

The variance of the rating of consumer i who purchases the product, $E(\mu_i| \theta_i) \geq p$, is obtained by

$$Var(r_i \mid \theta_i \geq \alpha) = Var(q_i \mid \theta_i \geq \alpha) = E(q_i - E(q_i \mid \theta_i \geq \frac{\sigma_\theta(p - q^e) + \rho\sigma_q\mu_\theta}{\sigma_\theta + \rho\sigma_q}))^2$$

$$= E(q_i - q - \rho\frac{\sigma_q}{\sigma_\theta}(\theta_i - \mu_\theta) - \sqrt{1-\rho^2}\sigma_q\Lambda_0)^2$$

$$= E(q_i - q - \rho\frac{\sigma_q}{\sigma_\theta}(\theta_i - \mu_\theta))^2 - (1-\rho^2)\sigma_q^2\Lambda_0^2$$

$$= \int_{-\infty}^{\underline{\delta}+q^e+\rho\frac{\sigma_q}{\sigma_\theta}(\theta_i-\mu_\theta)} \frac{\frac{(q_i - q - \rho\frac{\sigma_q}{\sigma_\theta}(\theta_i - \mu_\theta))^2}{\sqrt{1-\rho^2}\sigma_q}\phi(\frac{q_i - q - \rho\frac{\sigma_q}{\sigma_\theta}(\theta_i - \mu_\theta)}{\sqrt{1-\rho^2}\sigma_q})}{\Phi(\frac{\underline{\delta}+q^e-q}{\sqrt{1-\rho^2}\sigma_q}) + (1-\Phi(\frac{\overline{\delta}+q^e-q}{\sqrt{1-\rho^2}\sigma_q}))}dq_i + \quad \text{[D5]}$$

$$\int_{\overline{\delta}+q^e+\rho\frac{\sigma_q}{\sigma_\theta}(\theta_i-\mu_\theta)}^{+\infty} \frac{\frac{(q_i - q - \rho\frac{\sigma_q}{\sigma_\theta}(\theta_i - \mu_\theta))^2}{\sqrt{1-\rho^2}\sigma_q}\phi(\frac{q_i - q - \rho\frac{\sigma_q}{\sigma_\theta}(\theta_i - \mu_\theta)}{\sqrt{1-\rho^2}\sigma_q})}{\Phi(\frac{\underline{\delta}+q^e-q}{\sqrt{1-\rho^2}\sigma_q}) + (1-\Phi(\frac{\overline{\delta}+q^e-q}{\sqrt{1-\rho^2}\sigma_q}))}dq_i - (1-\rho^2)\sigma_q^2\Lambda_0^2$$

$$= (1-\rho^2)\sigma_q^2(1 + \Lambda_1 - \Lambda_0^2)$$

Where:

$$\Lambda_1 = \frac{\frac{\overline{\delta}+q^e-q}{\sqrt{1-\rho^2}\sigma_q}\phi(\frac{\overline{\delta}+q^e-q}{\sqrt{1-\rho^2}\sigma_q}) - \frac{\underline{\delta}+q^e-q}{\sqrt{1-\rho^2}\sigma_q}\phi(\frac{\underline{\delta}+q^e-q}{\sqrt{1-\rho^2}\sigma_q})}{\Phi(\frac{\underline{\delta}+q^e-q}{\sqrt{1-\rho^2}\sigma_q}) + (1-\Phi(\frac{\overline{\delta}+q^e-q}{\sqrt{1-\rho^2}\sigma_q}))}$$

By the Law of Total Variance and Equation [D5]

$$\begin{aligned}\sigma_r^2 &= VarE\left(\left(r_1|\theta_i\right)\right) + E\left(Var\left(r_1|\theta_i\right)\right)\\ &= Var\left(q+\rho\frac{\sigma_q}{\sigma_\theta}(\theta_i-\mu_\theta)+\sqrt{1-\rho^2}\sigma_q\Lambda_0\right)+(1-\rho^2)\sigma_q^2(1+\Lambda_1-\Lambda_0^2)\\ &= \rho^2\sigma_q^2(1-\lambda(\alpha)(\lambda(\alpha)-\alpha))+(1-\rho^2)\sigma_q^2(1+\Lambda_1-\Lambda_0^2)\end{aligned} \qquad [D6]$$

Q.E.D.

Proof of Corollary 1:

Given p, q^e, and $\Lambda_0(q^e) = \dfrac{\phi\left(\frac{\bar{\delta}+(q^e-q)}{\sqrt{1-\rho^2}\sigma_q}\right)-\phi\left(\frac{\underline{\delta}+(q^e-q)}{\sqrt{1-\rho^2}\sigma_q}\right)}{\Phi\left(\frac{\underline{\delta}+(q^e-q)}{\sqrt{1-\rho^2}\sigma_q}\right)+\left(1-\Phi\left(\frac{\bar{\delta}+(q^e-q)}{\sqrt{1-\rho^2}\sigma_q}\right)\right)}$

$$\begin{aligned}
&sign\left(\frac{\partial E(r)}{\partial\bar{\delta}}\right) = sign\frac{\partial\Lambda_0\left(q^e\right)}{\partial\bar{\delta}}\\
&= sign\left(\frac{\partial\phi\left(\frac{\bar{\delta}+(q^e-q)}{\sqrt{1-\rho^2}\sigma_q}\right)}{\partial\bar{\delta}}\left[\Phi\left(\frac{\underline{\delta}+(q^e-q)}{\sqrt{1-\rho^2}\sigma_q}\right)+\left(1-\Phi\left(\frac{\bar{\delta}+(q^e-q)}{\sqrt{1-\rho^2}\sigma_q}\right)\right)\right]\right.\\
&\qquad\left.-\left[\Phi\left(\frac{\bar{\delta}+(q^e-q)}{\sqrt{1-\rho^2}\sigma_q}\right)-\phi\left(\frac{\underline{\delta}+(q^e-q)}{\sqrt{1-\rho^2}\sigma_q}\right)\right]\frac{\partial\left[\Phi\left(\frac{\underline{\delta}+(q^e-q)}{\sqrt{1-\rho^2}\sigma_q}\right)\right]+\left(1-\Phi\left(\frac{\bar{\delta}+(q^e-q)}{\sqrt{1-\rho^2}\sigma_q}\right)\right)}{\partial\bar{\delta}}\right)\\
&= sign\left(\phi'\left(\frac{\bar{\delta}+(q^e-q)}{\sqrt{1-\rho^2}\sigma_q}\right)\left[\Phi\left(\frac{\underline{\delta}+(q^e-q)}{\sqrt{1-\rho^2}\sigma_q}\right)+\left(1-\Phi\left(\frac{\bar{\delta}+(q^e-q)}{\sqrt{1-\rho^2}\sigma_q}\right)\right)\right]\right.\\
&\qquad\left.+\phi\left(\frac{\bar{\delta}+(q^e-q)}{\sqrt{1-\rho^2}\sigma_q}\right)\left[\phi\left(\frac{\bar{\delta}+(q^e-q)}{\sqrt{1-\rho^2}\sigma_q}\right)-\phi\left(\frac{\underline{\delta}+(q^e-q)}{\sqrt{1-\rho^2}\sigma_q}\right)\right]\right)\\
&= sign\left(-\phi\left(\frac{\bar{\delta}+(q^e-q)}{\sqrt{1-\rho^2}\sigma_q}\right)\left(\frac{\bar{\delta}+(q^e-q)}{\sqrt{1-\rho^2}\sigma_q}\right)\left[\Phi\left(\frac{\underline{\delta}+(q^e-q)}{\sqrt{1-\rho^2}\sigma_q}\right)+\left(1-\Phi\left(\frac{\bar{\delta}+(q^e-q)}{\sqrt{1-\rho^2}\sigma_q}\right)\right)\right]\right.\\
&\qquad\left.+\phi\left(\frac{\bar{\delta}+(q^e-q)}{\sqrt{1-\rho^2}\sigma_q}\right)\left[\phi\left(\frac{\bar{\delta}+(q^e-q)}{\sqrt{1-\rho^2}\sigma_q}\right)-\phi\left(\frac{\underline{\delta}+(q^e-q)}{\sqrt{1-\rho^2}\sigma_q}\right)\right]\right)\\
&= sign\left(\phi\left(\frac{\bar{\delta}+(q^e-q)}{\sqrt{1-\rho^2}\sigma_q}\right)-\phi\left(\frac{\underline{\delta}+(q^e-q)}{\sqrt{1-\rho^2}\sigma_q}\right)\right.\\
&\qquad\left.-\left(\frac{\bar{\delta}+(q^e-q)}{\sqrt{1-\rho^2}\sigma_q}\right)\left[\Phi\left(\frac{\underline{\delta}+(q^e-q)}{\sqrt{1-\rho^2}\sigma_q}\right)+\left(1-\Phi\left(\frac{\bar{\delta}+(q^e-q)}{\sqrt{1-\rho^2}\sigma_q}\right)\right)\right]\right)
\end{aligned}$$

Thus the sign of $\frac{\partial E(r)}{\partial\bar{\delta}}$ depends on the comparison of the $\Lambda_0\left(q^e\right)$ and $\frac{\bar{\delta}+(q^e-q)}{\sqrt{1-\rho^2}\sigma_q}$. When $\frac{\bar{\delta}+(q^e-q)}{\sqrt{1-\rho^2}\sigma_q} < \Lambda_0\left(q^e\right)$, that is, $\bar{\delta}+\left(q^e-q\right) < \Lambda_0\left(q^e\right)\sqrt{1-\rho^2}\sigma_q$, the last expression is positive, $\frac{\partial E(r)}{\partial\bar{\delta}} > 0$. That is, the expected rating increases with $\bar{\delta}$ and vice versa.

Similar to the above results, the sign of $\frac{\partial E(r)}{\partial \underline{\delta}}$ depends on the sign of $\Lambda_0\left(q^e\right)$.

$$
sign\left(\frac{\partial E(r)}{\partial \underline{\delta}}\right) = sign\left(\frac{\partial \Lambda_0\left(q^e\right)}{\partial \underline{\delta}}\right)
$$

$$
= sign\left(-\frac{\partial \phi\left(\frac{\underline{\delta}+\left(q^e-q\right)}{\sqrt{1-\rho^2}\sigma_q}\right)}{\partial \underline{\delta}} \left[\Phi\left(\frac{\underline{\delta}+\left(q^e-q\right)}{\sqrt{1-\rho^2}\sigma_q}\right) + \left(1-\Phi\left(\frac{\bar{\delta}+\left(q^e-q\right)}{\sqrt{1-\rho^2}\sigma_q}\right)\right)\right] - \left[\phi\left(\frac{\bar{\delta}+\left(q^e-q\right)}{\sqrt{1-\rho^2}\sigma_q}\right) \right.\right.
$$

$$
\left.\left. -\phi\left(\frac{\underline{\delta}+\left(q^e-q\right)}{\sqrt{1-\rho^2}\sigma_q}\right)\right] \frac{\partial\left[\Phi\Phi\left(\frac{\underline{\delta}+\left(q^e-q\right)}{\sqrt{1-\rho^2}\sigma_q}\right)+\left(1-\Phi\left(\frac{\bar{\delta}+\left(q^e-q\right)}{\sqrt{1-\rho^2}\sigma_q}\right)\right)\right]}{\partial \underline{\delta}} \right)
$$

$$
= sign\left(-\phi'\left(\frac{\underline{\delta}+\left(q^e-q\right)}{\sqrt{1-\rho^2}\sigma_q}\right)\left[\Phi\left(\frac{\underline{\delta}+\left(q^e-q\right)}{\sqrt{1-\rho^2}\sigma_q}\right) + \left(1-\Phi\left(\frac{\bar{\delta}+\left(q^e-q\right)}{\sqrt{1-\rho^2}\sigma_q}\right)\right)\right]\right.
$$

$$
\left. -\phi\left(\frac{\underline{\delta}+\left(q^e-q\right)}{\sqrt{1-\rho^2}\sigma_q}\right)\left[\Phi\left(\frac{\bar{\delta}+\left(q^e-q\right)}{\sqrt{1-\rho^2}\sigma_q}\right) - \phi\left(\frac{\underline{\delta}+\left(q^e-q\right)}{\sqrt{1-\rho^2}\sigma_q}\right)\right]\right)
$$

$$
= sign\left(\left(\frac{\underline{\delta}+\left(q^e-q\right)}{\sqrt{1-\rho^2}\sigma_q}\right)\left[\Phi\left(\frac{\underline{\delta}+\left(q^e-q\right)}{\sqrt{1-\rho^2}\sigma_q}\right) + \left(1-\Phi\left(\frac{\bar{\delta}+\left(q^e-q\right)}{\sqrt{1-\rho^2}\sigma_q}\right)\right)\right] - \left[\phi\left(\frac{\bar{\delta}+\left(q^e-q\right)}{\sqrt{1-\rho^2}\sigma_q}\right)\right.\right.
$$

$$
\left.\left. -\phi\left(\frac{\underline{\delta}+\left(q^e-q\right)}{\sqrt{1-\rho^2}\sigma_q}\right)\right]\right)
$$

When $\frac{\bar{\delta}+\left(q^e-q\right)}{\sqrt{1-\rho^2}\sigma_q} < \Lambda_0\left(q^e\right)$, that is, $\bar{\delta}+\left(q^e-q\right) < \Lambda_0\left(q^e\right)\sqrt{1-\rho^2}\sigma_q$, the expected rating decreases with $\underline{\delta}$, $\frac{\partial E(r)}{\partial \underline{\delta}} < 0$, and vice versa.
Q.E.D.

Proof to Corollary 2:

(i) By Proposition 1, the expected consumer rating in a single period is $q+\rho\sigma_q\lambda\left(\alpha\left(p,q^e\right)\right)+\sqrt{1-\rho^2}\sigma_q\Lambda_0\left(q^e\right)-p$ and variance

$$
\rho^2\sigma_q^2\left(1-\lambda\left(\alpha\left(p,q^e\right)\right)\left(\lambda\left(\alpha\left(p,q^e\right)\right)-\alpha\left(p,q^e\right)\right)+\left(1-\rho^2\right)\sigma_q^2\left(1+\Lambda_1\left(q^e\right)-\Lambda_0^2\left(q^e\right)\right)\right).
$$

Since consumers in different periods are updating their quality beliefs based on Equation [10], $q_{t+1}^{2} = \omega q + (1-\omega)E\left(r_t\right)$. If consumers can fully overcome both types of biases, then $q_{t+1}^e = q$. Plug into Proposition 1, then the mean of the rating series keeps the same over time: $q+\rho\sigma_q\lambda\left(\alpha(p,q)\right)+\sqrt{1-\rho^2}\sigma_q\Lambda_{-}-p$ and variance also does not change over time $\rho^2\sigma_q^2\left(1-\lambda\left(\alpha(p,q)\right)\left(\lambda\left(\alpha(p,q)\right)-\alpha(p,q)\right)\right)+\left(1-\rho^2\right)\sigma_q^2\left(1+\Lambda_1-\Lambda_0^2\right)$. Thus the rating series is stationary.

(ii) If consumers form quality expectation without the biases in the previous periods, $q_{t=1}^e = q$ and
$E\left(r_t\right) = q+\rho\sigma_q\lambda\left(\alpha(p,q)\right)+\sqrt{1-\rho^2}\sigma_q\Lambda_0-p$,
$Var\left(r_t\right) = \rho^2\sigma_q^2\left(1-\lambda\left(\alpha(p,q)\right)\left(\lambda\left(\alpha(,q)\right)-\alpha(p,q)\right)\right)+\left(1-\rho^2\right)\sigma_q^2\left(1+\Lambda_1-\Lambda_0^2\right)$. The correlation between ratings in different time periods $corr(r_t, r_{t+1})$ thus rating series are independent.
Q.E.D.

References

Aigner, D. J. 1971. *Basic Econometrics*, Upper Saddle River, NJ: Prentice Hall
Anderson, E. W. 1998. "Customer Satisfaction and Word of Mouth," *Journal of Service Research* (1:1), pp. 5-17.
Anderson, T. W. 2005. "Origins of the Limited Information Maximum Likelihood and Two-Stage Least Squares Estimators," *Journal of Econometrics* (127), pp. 1-16.
Chakravarti, I. M., Laha, R. G., and Roy, J. 1967. *Handbook of Methods of Applied Statistics*, New York: John Wiley and Sons.
Chen, P-Y., Wu, S-Y., and Yoon, J. 2004. "The Impact of Online Recommendations and Consumer Feedback on Sales," in *Proceedings of the 25th International Conference on Information Systems*, Washington, D.C., pp. 711-724.
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), pp. 579-589
Chevalier, J., and Mayzlin, D. 2006. "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research* (43:3), pp. 345-354.
Chintagunta, P., Erdem, T., Rossi, P. E., and Wedel, M. 2006. "Structural Modeling in Marketing: Review and Assessment," *Marketing Science* (25:6), pp. 604-616.
Clemons, E. K., Gao, G., and Hitt, L. M. 2006. "When Online Reviews Meet Hyper Differentiation: A Study of the Craft Beer Industry," *Journal of Management Information Systems* (23:2), pp. 149-171.
Davidson, R., and MacKinnon, J. G. 1993. *Estimation and Inference in Econometrics*, Oxford, UK: Oxford University Press.
Diebold, F., and Kilian, L. 2001. "Measuring Predictability: Theory and Macroeconomic Applications," *Journal of Applied Econometrics* (16:6), pp. 657-669.
Hartigan, J. A., and Hartigan, P. M. 1985. "The DIP Test of Unimodality," *Annals of Statistics* (13:1), pp. 70-84.
Henderson, R., Diggle, P. J., and Dobson, A. 2000. "Joint Modeling of Longitudinal Measurements and Event Time Data," *Biostatistics* (1), pp. 465-480.
Muller, D. W., and Sawitzki, G. 1991. "Excess Mass Estimates and Tests of Multimodality," *Journal of the American Statistical Association* (86), pp. 738-746.
Silverman, B. W. 1981. "Using Kernel Density Estimates to Investigate Multimodality," *Journal of the Royal Statistical Society* (43:1), pp. 97-99.
Stephens, M. A. 1974. "EDF Statistics for Goodness of Fit and Some Comparisons," *Journal of the American Statistical Association* (69), pp. 730-737.
Thode Jr., H. C. 2002. *Testing for Normality*, New York: Marcel Dekker.