Effect of Disconfirmation on Online Rating Behavior: A Dynamic Analysis

Yi-Chun "Chad" Ho • Yong Tan

University of Washington, Seattle, WA 98105

Junjie Wu

Beihang University, Beijing, China

Abstract

This study attempts to answer a basic question: Under what conditions are consumers more likely to share their product evaluation online? Unlike prior research which assumes that the influence of existing ratings by others on subsequent ratings occurs in the post-purchase stage only, we postulate that such impact may take place when prospective buyers receive review information in the pre-purchase stage. To investigate the dynamics of consumers' rating behavior, we model how an individual's perception of system credibility evolves over time, and how such perception dynamically affects her decision of whether to post an online rating after product experience.

Our estimation results show that an online consumer's posting decision is driven by *quality disconfirmation*, a term which captures the discrepancy between her expected and realized quality obtained from the same product. That is, consumers are more likely to express their product opinions when the ex-post evaluation further deviates from the ex-ante expectation. Moreover, the effect of quality disconfirmation on posting behavior is moderated by rating environments where the focal consumer is exposed. Our results also show that online consumers tend to become less active in review contribution as they perceive the review system to be more reliable over time. Through simulations, we illustrate 1) the dissimilar reporting propensity across different levels of product evaluation; 2) the evolution of online product ratings over time; and 3) the effect of review manipulation on subsequent rating entries. This research contributes to the literature by providing a comprehensive understanding on the formation of online opinions based on a rich economic reasoning.

Keywords: online product reviews, consumer rating behavior, disconfirmation, structural modeling, hierarchical Bayes, Bayesian learning

1. Introduction

It has been well recognized that online product reviews and ratings have a substantial impact on consumers' pre-purchase decisions. According to surveys, 82% of the respondents agree that online reviews directly influence their decisions on what product to buy (Deloitte 2007) and over 75% of them consider recommendations from those who have experienced the product the most credible information source (Nielson 2007). Through observing the product evaluation shared by peer consumers, a prospective buyer can reduce the product uncertainty and make a more informed purchase, leading to higher satisfaction and lower merchandise return rates (PowerReviews 2010). As a result, the prevalence of online reviews enhances the information transparency and therefore improves the efficiency of online marketplace.

An online product review platform is an information system which facilities information exchange among its users. Users of an online rating system can be categorized into two groups based on how they interact with the system. The first type of users is review reader (information receiver) who gathers product information stored in various formats, such as numeric ratings, textual contents, multi-media files, etc. After purchasing and experiencing the product, the review reader has an opportunity to become a review poster (information provider) by sharing her own evaluation about the product. Interestingly, a majority of system users are observed to behave as "lurkers", which refers to those who read reviews only but do not post. While how observed ratings can be related to product performance has been studied in a variety of contexts, lesser attention is paid to the underlying factors that induce posting behavior from consumers. Moreover, the distribution of product ratings[1] is commonly found to have a right-skewed U-shape or J-shape across various contexts and platforms (McGlohon et al. 2010). Although researchers have shown that individuals with extreme experiences tend to be more vocal from a statistical approach (Anderson 1998; Dellarocas and Narayan 2006), research to date has not provided an economic-grounded explanation to such observation in the information systems literature. In this study, we empirically examine online consumers' product rating behavior and attempt to explain above-mentioned phenomena commonly observed across online rating systems.

Motivated by the dual roles consumers can play when interacting with the system, we model review-posting behavior from two novel aspects. First, research has shown that, when making rating decision, an individual tends to first observe the opinions expressed by others and then adjust her own opinion accordingly. Such framework assumes that the previously posted ratings influence a focal individual's rating decisions in the post-purchase stage only, ignoring the fact that review posters are often review readers as well. In this paper, we posit that the impact of antecedent ratings may actually take place when

[1] By *rating* or *score* we mean the numeric value indicating overall product satisfaction experienced by consumers.

the focal individual observes the rating signal in the pre-purchase stage. The general rationale is that before making a purchase, a consumer forms the *expected quality* of the product based on the review signal she observes and the intrinsic perception of the system credibility she has. Upon product consumption, the consumer obtains the *realized quality* of the product and encounters *quality disconfirmation*, referring to as the deviation between her expected and realized quality obtained from the same product. We hypothesize that the realization of quality disconfirmation may have a direct impact on consumers' decision of whether or not to post a product rating and leave it as an empirical question.

The second novelty of this research is that we allow the underlying mechanism governing consumers' rating behavior to be dynamic. In particular, we consider that at a given time, an individual possesses a *perception of the system credibility*, a subjective attitude underlying how she interprets rating signals she observes. Such perception would evolve over time, depending on quality disconfirmation realized from each product experience. In this paper, we also investigate whether an online rater's posting behavior is dynamically affected by her perceived system credibility.

The richness of our data allows us to directly observe online consumers' decision of "whether to post" (posting vs. lurking) as well as the decision of "what to rate" in the post-consumption stage. To model the interdependence between two decisions, we adopt a Heckman selection model (Heckman 1979) to account for the raters' posting decision in the first stage, and generalize it to reflect the discrete outcomes observed in the second stage. To the best of our knowledge, this is the first research that studies how consumers' rating behavior evolves over time at the individual level. Using a panel data set consisting of each individual consumer's complete purchasing and rating activities taken place on an e-commerce site, we attempt to provide a more comprehensive understanding on the formation of online product opinions based on a richer economic reasoning.

Our estimation results show that a consumer's review-posting behavior is, at least partially, driven by the magnitude of quality disconfirmation she encounters, perhaps due to an intention to "correct" the review signal. This finding provides an economic explanation to the commonly observed U-shaped distribution of online ratings. Moreover, the impact of quality disconfirmation on review contribution is further moderated by the rating environment where a focal consumer is exposed. In particular, the disconfirmation effect is weakened by the level of dissension among posted ratings; however, it is intensified by the total number of submitted reviews. We also find that consumers tend to be more vocal if they perceive the review system to be noisy, relatively to the scenario where the system is perceived highly reliable.

To highlight the significance of the disconfirmation effect identified in this research, we perform a series of simulations to better understand consumers' posting behavior and evolution of online product ratings. First, we recover the unobserved ratings for purchase occasions that do not lead to a review entry.

Our simulation results show that, at the population level, the relationship between review posting percentage and realized product quality is best described as a left-skewed U shape. That is, negative product experiences strongly induce posting behavior. Our second simulation demonstrates that the average of posted product ratings will converge to the true quality in the long run. Although other researchers have explained why average product ratings decline over time (Li and Hitt 2008; Moe and Schweidel 2012), we believe that the disconfirmation effect provides an alternative explanation as it can explain other different evolution patterns observed in practice.

The rest of this paper is organized as follows. In Section 2, we briefly review the relevant literature and position our work. Section 3 presents our conceptual framework and Section 4 describes the data. Our empirical model is developed in Section 5. Section 6 discusses estimation strategy, model fit and model comparison. In Section 7, we present our estimation results and discuss the associated insights. We also provide robustness checks and conduct a series of simulations to highlight the significance of our main findings. Concluding remarks and future research directions are provided in Section 8.

2. Literature Review

The stream of literature studying why consumers engage in post-purchase WOM is most related to this research. Using survey data, researchers investigate this question from a motivational point of view. In an offline setting, consumers' desire for altruism, product involvement and self-enhancement are main factors leading to positive WOM, whereas consumers spread negative WOM usually for altruism, anxiety reduction and vengeance purpose (Sundaram et al. 1998). With a similar approach, Hennig-Thurau et al. (2004) conclude that social benefits, economic incentives, concern for others, and extraversion are the primary motives for consumer expressing their product experiences on the Internet.

Based on quantitative methods, an extensive literature has also been developed to identify what drives consumers to share their product or service experiences in the absence of explicit reward mechanisms. Dellarocas et al. (2004) examine the rating behavior on an electronic trading platform. They find that such voluntary behavior is driven by the expectation of reciprocal behavior, meaning that a trader evaluates her trading partner in order to solicit feedback from the other party. Shen et al. (2013) look at the review posting behavior by consumers from a strategic perspective. Using the book review data from online book sellers, they conclude that online reviewers are more prone to rate popular but less crowded books in order to gain attention but reduce competition for attention at the same time. They also conclude that reviewers with high reputation costs are more likely to adopt an imitation strategy by posting ratings conforming to the community consensus. Factors affecting the level of WOM have also been studied at the population level. Rather than taking the common conception of the level of WOM activity as a monotonic function of customer satisfaction, Anderson (1998) discovers that the relationship between them exhibits a U-shaped pattern – customers are more likely to engage in WOM when they are either extremely satisfied or extremely dissatisfied. Using the data from a movie website, Dellarocas and Narayan (2006) also identify a similar association between observed rating density and perceived movie quality. Along this line, Dellarocas et al. (2010) further suggest that movie goers are more prone to post review for the most or least popular movies in terms of box office revenues.

There is an emerging literature stream examining how existing ratings affect subsequent ones. In a lab setting, Schlosser (2005) identifies the "self-presentational" phenomenon in which a review poster strategically adjusts her product ratings downwards after observing negative opinions by others in order to present herself as intelligent. She also finds that a consumer would make her opinions more balanced if the opinions from the crowd demonstrate a high level of dissention. Li and Hitt (2008) conclude that predominant declining trend of book ratings can be attributed to consumers' "self-selection" bias, meaning that early buyers have higher perceived quality and therefore tend to give higher ratings than do later buyers. Using reviews posted on an online retailer of bath, fragrance and home products, Moe and Schweidel (2012) show that consumers' rating behavior is influenced by rating environment they are exposed to. In particular, they find that a consumer is more inclined to share her experience when the posted ratings are more positive as measured by high valance and high volume. In addition, active posters are found to be more negative and exhibit differentiation behavior, which is consistent with the self-presentational strategy suggested by Schlosser (2005). Lee et al. (2013) distinguish prior ratings by the friends from those by the strangers and investigate whether these two measurements have different impact on a focal individual's opinions. They find that friends' opinions always induce herding and the presence of social networking dampens the impact of opinions from the crowd.

In terms of research context and methodology, our work is most relevant to the following two papers that explicitly model online reviewers' rating behavior at the individual level. Ying et al. (2006) argue that knowledge that a product was rated (selected) should affect analyst's prediction of that rating. After accounting for reviewers' selection problem, they show that the performance of movie recommendation can be substantially improved. Moe and Schweidel (2012) further generalize Ying's model by flexibly linking a reviewer's decision of "whether to rate" and that of "what to rate".[2] Research studying consumers' review posting decision naturally leads to another stream of literature which investigates whether consumer-generated reviews can represent true product quality. This particular literature stream can be further categorized into two sub-streams, depending on whether publicly available reviews are completely generated by consumers or partially manipulated by firms. Following the notion that online ratings are truly truthful, Hu et al. (2006) discuss whether the mean of posted ratings can represent true

[2] We provide a detailed discussion on the difference between our work and Moe and Schweidel (2012) in Section 5.

product quality. In particular, they develop an analytical model which assumes that a consumer would post online reviews only when the level of her satisfaction is either above or below a "brag-and-moan" interval; she would otherwise behave as a lurker. Based on this assumption, they show that the average rating can serve as an unbiased signal if and only if two bounds of the interval are equal or symmetrically deviate from the true quality. On the other hand, Dellarocas (2006) assumes that the observed ratings may not be fully trustful and could be strategically boosted by firms. He demonstrates that inflated reviews are still informative since the firm producing high-quality products benefits the most through manipulation.

In this paper, we attempt to contribute to the online review literature by looking at the relationship between posted ratings and subsequent ones from a novel perspective. In contrast to extant work that focus on how others' opinions influence a consumer's rating behavior in the post-purchase stage, we postulate that the interaction between them may take place when a prospective buyer reads reviews for information acquisition in the pre-purchase stage. Such modeling novelty allows us to develop a utility-based framework and examine the underlying mechanism driving consumers to voluntarily share their product experiences. In addition, we account for the dynamics of rating behavior by modeling how a raters' perception of system credibility evolves over time and how such perception affects her propensity to post dynamically.

3. Research Framework and Context

In this section, we introduce our conceptual framework and illustrate the underlying mechanism governing consumers' rating behavior[3]. According to the theory of buying decision, after consumers recognize the need for a certain product, consumer would gather information about the product in the pre-purchase stage. The emergence of online product rating system facilitates communication among consumers such that they can easily exchange their product experiences. In our research context, the e-commerce website displays the arithmetic mean and other aggregate statistics of all posted ratings at the top of a product's landing page.[4] In the pre-purchase stage, a prospective buyer observes the rating signal associated with the product she desires to purchase and formulates an expected quality. Upon product consumption, the consumer obtains realized quality of the product per how much she enjoys the product. In the post-purchase stage, the consumer has another choice to make. She could either become a review poster by expressing her own product opinions; or, she could simply remain silent and become a "lurker". The realized quality has been shown to have explanatory power in eWOM engagement in many contexts

[3] We use the term *rating* as a general expression of consumers' product review behavior, whereas the term *posting* as a decision-making process in which consumers decide whether to post a product review after product consumption.

[4] Like other typical e-commerce sites, the total number and the distribution of all posted ratings are also publicly available on product landing pages.

such as movies (Dellarocas and Narayan 2006), bath, fragrance, and home products (Moe and Schweidel 2012), etc. Along this research stream, we further speculate that consumers' posting decision may also be influenced by "*quality disconfirmation*", defined as the discrepancy between one's expected and realized quality acquired from the same product (see Figure 1).

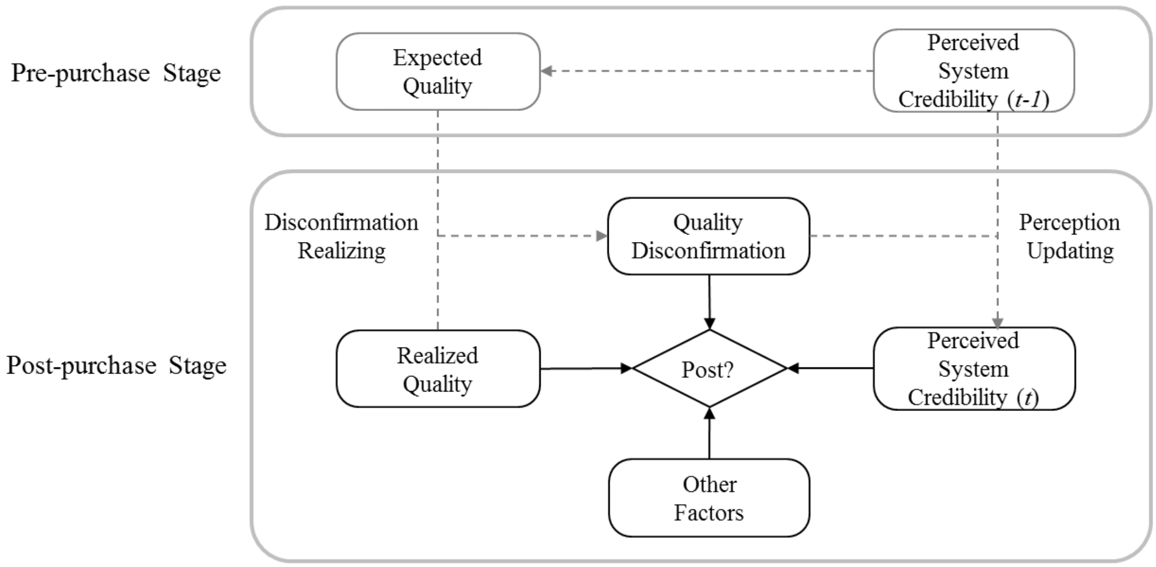
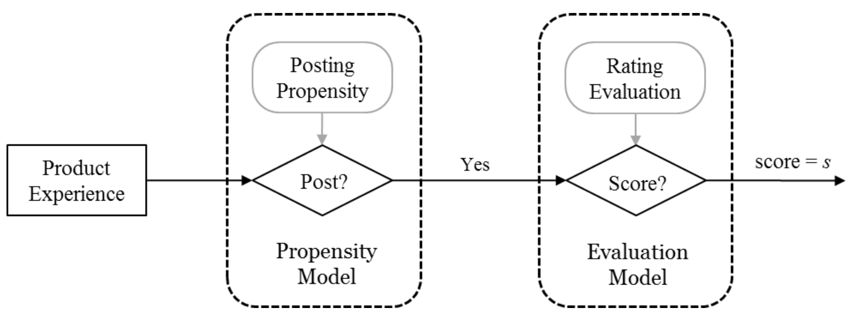


Figure 1. Conceptual Framework of Online Rating Behavior

The effect of quality disconfirmation is considered to have a short-term impulse on the rating decision. While survey-based research (Hennig-Thurau et al. 2004; Sundaram et al. 1998) has identified several motivations for consumers engaging in WOM from a *normative* perspective, our utility-oriented framework provides *positive* explanations on why people share their evaluation after product consumption. For example, concerns for other are found to be one of the most important motives. In our framework, the motive "concern for others" can be explained as an altruistic behavior driven by consumers' quality disconfirmation. If an individual perceives that the average rating overrates (underrates) the true product quality, she would have incentive to "correct" the rating signal by reporting her own evaluation and bringing the mean score downwards (upwards). Similarly, other motivations such as anxiety reduction and vengeance can also be triggered by negative disconfirmation. If the hypothesized disconfirmation effect does exist, we are also interested in whether such effect is moderated by the rating environments to which online consumers are exposed.

Since a review poster is often a review reader as well, we posit that a consumer's perceived system credibility may also affect her willingness to interact with the system. In particular, we model consumers' perception of the system credibility is composed of two components: biasness and reliability. Consider a consumer who possesses a certain perception of system credibility. Before purchasing a product, she first observes the previously posted ratings, which provides some information about the product quality. Based on this rating signal, the consumer forms the expected quality of the product, taking into account her own perception of the system biasedness and reliability. For example, if the consumer perceives the system to be unbiased, she would formulate an expected quality based on the rating signal she observes. If she perceives the review system to be positively biased (i.e. the rating signal inflates the product quality), she would adjust her prior expectation downwards, and vice versa. Having experienced the products, the consumer encounters quality disconfirmation and then uses this private information to update her beliefs in system biasedness and reliability. If the rating signals consistently portray her own product evaluation, she would perceive the system to be more reliable. On the other hand, if the rating signal consistently deviates from her product evaluation to a large degree, she would perceive the system to be noisy. Since the perceived system credibility is individual-specific and time-variant, we are able to investigate whether such perception would affect one's review-posting decisions over time. While credibility has been investigated in different applications such as how it determines the persuasiveness of communication (Chaiken 1980), its effect on online opinions expression still remains unstudied. Incorporating the credibility component allows us to model consumers' rating behavior in a dynamic fashion.

We assume that an individual's rating behavior is governed by a two-stage mechanism[5] (Ying et al. 2006), illustrated in Figure 2. In the stage 1, she decides whether or not to express her opinion by posting a product rating (posting decision). If she decides to leave a review, at the same time, she also chooses what rating to give (rating decision) in the stage 2. While a large body of literature implicitly assumes review incidences occur haphazardly, we believe this assumption is unrealistic. In particular, we model that a consumer's posting decision is determined by an attitude which we call *posting propensity*. The posting propensity is composed of four components: an individual's realized quality, quality disconfirmation, perception of system reliability and some other factors such as product price and characteristics of posting ratings. If an individual's posting propensity exceeds a certain threshold, she would share her own experience through leaving an online product review. Meanwhile, she would also need to decide what score to submit according to her *rating evaluation*, an attitude mainly determined by her overall evaluation of product experience.



[5] We call it a two-stage mechanism for the sake of modeling. Consumers can make two decisions simultaneously.

Figure 2. Consumers' Online Rating Decisions

4. Data

The data for this study is provided by an online e-commerce site which sells a variety of search goods, such as home and kitchen items, toys, electronics, etc. The data set contains complete purchase history and review entries made by 1,000 heavy users who made at least 10 purchases each. The transactional record set consists of order-specific information such as the product name, price, order handling time, order submission date, the statistics of all posted ratings associated with the product when the order was placed.[6] The rating data set stores user-generated reviews in a typical format, including a review title, a review body, submission date, and an overall product rating on a discrete 5-star scale, with 5 being the best. The data spans from January 2008 to October 2011. During this period of time, the site had not gone through any policy or system design change that could influence consumers' rating behavior. It did not implement any monetary and reviewers' reputation mechanisms that could possibly encourage posting. Participation in product review can be considered voluntary and self-driven.

It is worth noting that our data set is collected at the individual level. This nature distinguishes our work from others that commonly use review data at aggregate (product) level. The complete logs of purchasing and rating activities allow us to discern whether a purchase occasion leads to a rating entry. The time stamps associated with each activity also enables us to calibrate online rating behavior from a dynamic perspective. To identify the realized quality and consumers' learning dynamics, we exclude users who did not post any review. The resulting data set contains 361 panelists who made 37,209 purchase transactions and 2,257 review entries. We will use this dataset for model estimation. We supplement our primary data set by collecting the aggregate review information from the same platform in June 2013. This supplementary data provides the rating environments for each product and we will utilize this piece of information to proxy true product quality in our empirical model.

Before beginning the model estimation, we briefly present some observations and address the potential issues identified in the data. Table 1 reports the descriptive statistics of rating outcomes observed in our final data set. At the population level, the posting percentage is 6.06% and the mean of all posted ratings is about 3.83. Figure 3 plots the frequency of each of 5 discrete ratings. The distribution roughly follows a right-skewed U shape, which is commonly found across various rating platforms (Dellarocas and Narayan 2006; McGlohon et al. 2010). Such unique shape implies that consumers are more prone to express opinions when they are either very satisfied (represented by a high score of 4 or 5) or very dissatisfied (represented by a low score of 1 or 2), identified by Anderson (1998). Some control

[6] Like other online merchants, the user-generated textual reviews and numeric ratings are publicly available on the e-commerce site. Using the time stamp of purchase orders, we are able to recover the rating environment (characterized by the valence, volume and variance of posted ratings) at any time in the past.

variables, such as price and order handling time, exhibit a long-tail property (see Figure 4a). To deal with the over-dispersion issue, we take logarithmic transformation on them. The log-transformed variables appear to be normally distributed (see Figure 4b).

Table 1. Summary of Outcome Variables

Notations	Description	Mean	St. Dev.	Min	Max
Y_{ijt}	A dummy indicating if a purchase occasion leads to a review posting	0.061	0.239	0	1
Z_{ijt}	Observed rating scores	3.826	1.200	1	5

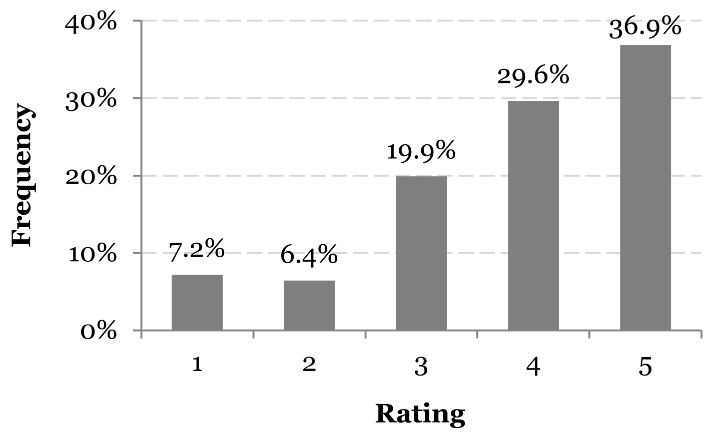


Figure 3. Frequency of Observed Rating Scores

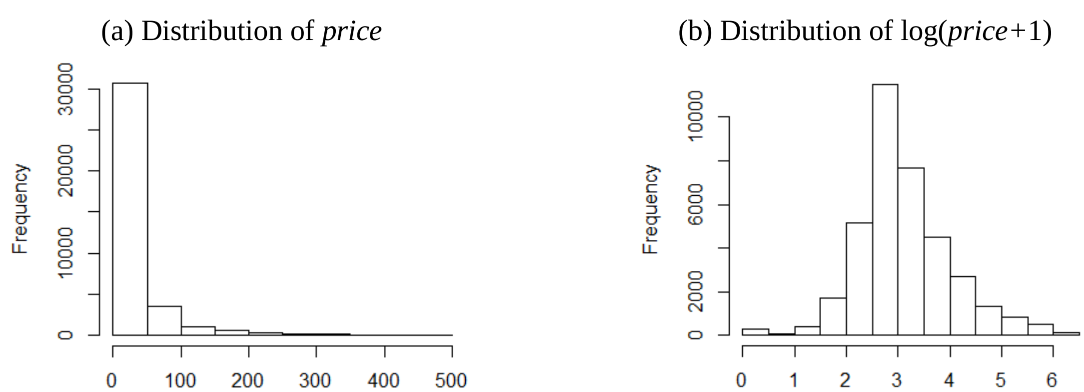


Figure 4. Over-dispersion vs. Logarithm Transformation

5. The Model

The paper which is closest to this study, both in terms of the research content and model itself, is Moe and Schweidel (2012) (henceforth MS). We develop our model based on MS which, in turn, is a generalization of Ying et al. (2006) who first model online consumers' rating behavior using a two-stage mechanism. Despite several similarities, this paper differs from MS in the following aspects.

Research question. The goal of MS is to study how online reviewers adjust their rating decision according to the rating environment they are exposed to. The impact of previously posted ratings on a focal individual's rating decision is assumed to take place in the post-purchase stage only. Instead, we speculate that such impact may occur when the individual consumes review information in the pre-purchase stage. The novelty of our research is that we explicitly model consumers' quality disconfirmation and their perception of system credibility in an effort to better understand what could possibly induce posting behavior from consumers.

Data. The data set used in MS includes a total of 3,681 product ratings provided by 2,436 unique reviewers; an individual on average posts 1.5 reviews only. With such limited data points, the estimates of individual-specific parameters in MS may not be able to precisely capture the intrinsic characteristics of each individual. On the other hand, the final data set we use for model estimation contains 361 individuals who made a total of 37,209 purchasing transactions and contributed 2,257 product reviews. The data richness and completeness allow us to more accurately gauge the heterogeneity cross individuals. Moreover, we can exactly discern whether or not a consumer posts an online rating at all after experiencing the product. With such data advantage, we can directly observe the posting decision for each purchase occasion without incorporating a latent purchase model.

Model specification. In MS, product quality is estimated using a zero-mean random effect. To utilize publicly available review information, we collect and use the long-term average rating as a proxy for true quality. We also explicitly model the interdependence between raters' decisions of whether to post and what to rate.

Covariates specification. MS apply a factor analysis and use the resulting factors as independent variables to explain reviewers' rating behavior. In our framework, we construct quality disconfirmation and perception of system reliability as the main explanatory variables. In addition, we also include several control variables such as product price, order handling time and aggregate statistics of expressed opinions by others to better model consumers' posting decisions.

Following the conceptual framework introduced earlier, we develop a two-stage selection model to explicitly model the interdependence between two rating decisions. The general modeling context is that in the pre-purchase stage, a consumer first formulates an expected quality right before she purchases the product. Such ex-ante expectation is formed based on the product ratings posted by peer consumers and the focal consumer's perceived system credibility. Upon product experience, she obtains the realized (ex-post) quality in the post-purchase stage. With expected and realized quality in mind, the individual encounters quality disconfirmation and use this private information to update her perception of the system credibility. Driven by quality disconfirmation and others factors, the consumer faces a two-stage decision with the first decision being whether to contribute a product rating; if she decides to post a rating, she also needs to choose what rating to leave[7]. The proposed model is described in the following order: 1) formulation of quality disconfirmation; 2) evolution of perceived system credibility; 3) consumers' decisions of whether to post and what to rate; and 4) interdependence between two rating decisions.

5.1. Formulation of Quality Disconfirmation

Expected Quality. In the pre-purchase stage, consumers formulate expected quality of the product based on information gathered from different sources. In our context, the e-commerce website displays the aggregate statistics of posted ratings[8] (e.g., arithmetic mean), on the top of product landing pages. Given the significance of online product ratings, it is reasonable to assume that consumers' expected quality of the product is largely influenced by this piece of information.[9] While posted reviews provide objective information about product quality, how such information is perceived could be subjective and heterogeneous across individuals.

Consider an individual *i* who is about to purchase a product *j* at time *t*. We assume the expected quality associated with this purchase occasion, $\hat{Q}_{ijt}$, to follow a normal distribution:

$$\hat{Q}_{ijt} \sim N\left(\bar{R}_{jt} + \delta_{it},\ \left(\tau_{it} \cdot \mathrm{T}_{jt}\right)^{-1}\right), \tag{1}$$

where $\bar{R}_{jt}$ and T_{jt} denote the arithmetic mean[10] and precision[11] of all ratings for *j* posted prior to *t*, respectively. In this paper, we model that consumers' perception of system credibility is composed of two components: biasedness (δ_{it}) and reliability (τ_{it}). The system biasedness δ_{it} measures how biased the review system is perceived by the individual *i* at time *t* and how she adjusts such bias. When $\delta_{it} = 0$, the individual *i* believes that $\bar{R}_{jt}$ provides unbiased information about the quality, and therefore, she would formulate an expected product quality centered on $\bar{R}_{jt}$. When $\delta_{it} > 0$ ($\delta_{it} < 0$), she believe that the rating signal $\bar{R}_{jt}$ underrates (overrates) the true product quality. As a result, she would adjust the mean of her expectation upwards (downwards). The system reliability τ_{it} measures how reliable the system is

[7] Again, these two decision-making processes can happen simultaneously from a consumer's perspective.
[8] These aggregate statistics include the average, the total number and the distribution of previously posted ratings.
[9] Consumers' ex-ante perception of quality is also influenced by textual product reviews (Ghose et al. 2012). However, we cannot observe how many textual reviews each prospective buyer reads before making a purchase decision. Therefore, we use the mean of posted ratings to measure the overall opinions shared by others.
[10] According to a survey conducted by Lightspeed Research (Leggatt 2011), 72% of online shoppers expect to find customer reviews available on the website they are shopping at, while 47% seek them out on company websites and 43% on price comparison sites. Therefore, we assume that one's expected quality is, by large part, determined by $\bar{R}_{jt}$, the mean ratings observed on the same site.
[11] We use precision instead of variance because the former gives us a neater expression of belief updating.

perceived by i at time t. When τ_{it} is large (small), the individual i would perceive the review system to be reliable (noisy) such that her expected quality would be tightly (loosely) centered on its mean.

Realized Quality. One of the most challenging parts of our modeling task is to model the baseline quality of products. One approach is to model the latent quality of all products using a zero-mean random effect (Moe and Schweidel 2012). To best utilize the publicly available information and enhance the estimation reliability, we use the long-term average rating, $\tilde{R}_j$, as a proxy for the baseline product quality.[12,13] In a spirit of Moe and Schweidel 2012, we assume that an individual i obtains realized quality Q_{ijt} from product j purchased at time t:

$$Q_{ijt} = \tilde{R}_j + \lambda_{i0}\,, \tag{2}$$

where the parameter λ_{i0} allows for variation in realized quality across individuals.

Quality disconfirmation. Having developed the expected and realized quality, we are now able to formally define quality disconfirmation. We model quality disconfirmation as how far away an individual's realized quality deviates from her expected quality:

$$\Delta Q_{ijt} \equiv Q_{ijt} - \hat{Q}_{ijt}\,. \tag{3}$$

Plugging (1) and (2) into (3), we have:

$$\Delta Q_{ijt} \sim N\left(\Delta\bar{Q}_{ijt},\ \left(\tau_{it} \cdot \mathrm{T}_{jt}\right)^{-1}\right), \tag{4}$$

$$\text{where}\ \ \Delta\bar{Q}_{ijt} = (\tilde{R}_j + \lambda_{i0}) - (\bar{R}_{jt} + \delta_{it})\,. \tag{5}$$

In this paper, we posit that the occasion-specific variable ΔQ_{ijt} not only directly affects the individual i's posting decision associated with the *ijt*-th occasion; it also indirectly influences i's long-term posting behavior through the evolution of her perceived system credibility.

5.2. Evolution of the Perception of System Credibility

Next, we explain how consumers' perception of the system credibility evolves over time. We assume that at time t an individual i's prior beliefs of $\delta_{it} \mid \tau_{\mathrm{it}}$ and τ_{it} jointly follow a normal-gamma distribution[14]:

---

[12] We justify our use of long-term average ratings as a proxy for quality in following two aspects. From the perspective of the volume of ratings, the mean and minimal values of the total number of posted ratings at product level are 67 and 11, respectively. From the perspective of time horizon, the time we collected our long-term rating data set (June 2013) is 18 months after the latest purchase occasion observed in our primary data set (November 2011). It is reasonable to justify that our long-term average ratings are representative of opinions from a sufficiently large consumer base and have converged to the true quality.

[13] We perform a robustness check in Section 7.1.3 and the results provide supportive evidence that the long-term average rating well proxy the true product quality.

[14] In Bayesian learning framework, quality disconfirmation defined here follows a normal model with unknown mean and unknown variance. The conjugate prior for this model is a normal-gamma joint distribution.

$$\delta_{it} \mid \tilde{\tau} \sim \mathrm{N}\left(\bar{\delta}_{it},\ \left(\gamma_{it} \cdot \tilde{\tau}\right)^{-1}\right), \tag{6}$$

$$\tau_{it} \sim \Gamma(a_{it},\ b_{it}), \tag{7}$$

where γ_{it} is the precision parameter of a normal distribution, a_{it} is the shape parameter and b_{it} is the inverse scale parameter of a gamma distribution. Upon product experience, the individual encounters quality disconfirmation and use this private information to update her beliefs of δ_{it} and τ_{it} jointly. As a result, an individual's perceived system credibility will evolve over time, as she is involved in more purchasing and experiencing activities. Based on Bayes' rule, the individual i's posterior beliefs after receiving a disconfirmation signal associated with occasion t+1 are given by (DeGroot 1970):

$$\delta_{i,t+1} \mid \tilde{\tau} \sim \mathrm{N}\left(\bar{\delta}_{i,t+1},\ \left(\gamma_{i,t+1} \cdot \tilde{\tau}\right)^{-1}\right), \tag{8}$$

$$\tau_{i,t+1} \sim \Gamma\left(a_{i,t+1}, b_{i,t+1}\right), \tag{9}$$

where

$$\bar{\delta}_{i,t+1} = \bar{\delta}_{it} + D_{j,t+1} \cdot \frac{\mathrm{T}_{jt}}{\gamma_{it} + \mathrm{T}_{jt}} \cdot \Delta\bar{Q}_{ij,t+1}, \tag{10}$$

$$\gamma_{i,t+1} = \gamma_{it} + D_{j,t+1} \tag{11}$$

$$a_{i,t+1} = a_{i,t} + D_{j,t+1} \cdot \frac{1}{2}, \tag{12}$$

$$b_{i,t+1} = b_{i,t} + D_{j,t+1} \cdot \frac{1}{2} \frac{\gamma_{it} \cdot \mathrm{T}_{jt} \cdot \left(\Delta\bar{Q}_{ij,t+1}\right)^2}{\gamma_{it} + 1}, \tag{13}$$

and $D_{j,t+1}$ is a dummy variable indicating whether there is at least one rating for j posted prior to time t+1.

It is important at this time to point out how perception of the system credibility evolves over time. Consider an individual i who has beliefs of $\delta_{it} \mid \tau_{\mathrm{it}}$ and τ_{it} at time t. Suppose that i observes the rating signal for j and purchases j at time t+1. Upon product experience, she receives one signal of disconfirmation, $\Delta Q_{ij,t+1}$, and updates her beliefs based on this information. If there is no product rating available at time t+1 (i.e. $D_{j,t+1} = 0$), no belief updating would occur and individual i's beliefs remains unchanged. If the rating signal is available (i.e. $D_{j,t+1} = 1$), she would jointly update her beliefs in system biasedness and reliability. According to (10), the posterior mean $\bar{\delta}_{i,t+1}$ is the sum of prior mean and the weighted realization of disconfirmation, with the weight being $\mathrm{T}_{jt} / (\gamma_{it} + \mathrm{T}_{jt})$. The signal precision T_{jt} and prior precision γ_{it} can be interpreted as the strength of the disconfirmation signal and the individual i's confidence in her belief of system biasedness. Similarly, the extent of updating of the scale parameter $b_{i,t+1}$ is increasing in the magnitude of disconfirmation and the precision of the rating signal. When T_{jt} is small, a consumer anticipates the review signal to be noisy with a higher probability and therefore updates her belief in a relatively small amount, ceteris paribus.

What remain unspecified are the initial beliefs of each individual. The updating rule of the shape parameter implies that a_{i0} represents the richness of the individual *i*'s initial learning experience before receiving any disconfirmation signal. Since we observe complete purchase history for all individuals from our data set, we fix a_{i0} at a small number[15] across individuals because all consumers have the same and limited amount of learning experience with the e-commerce site at time 0. To allow for heterogeneity across individuals in their initial perceptions of system reliability, we assume $b_{i0} \sim \mathrm{N}(\bar{b}_0, \sigma_b^2)$, where $\bar{b}_0$ and σ_b^2 measures the mean effect and dispersion of b_{i0} across individuals, respectively. As for system biasedness, we assume $\delta_{it} = 0$ since it is reasonable to argue that, before receiving any disconfirmation signal, consumers would initially perceive the review system to be unbiased mainly due to lack of information. We also fix $\gamma_{i0} = 0.1$ to reflect the fact that consumers have an uninformative initial prior belief in δ_{it}.

It is worth noting that the belief updating mechanism specified in our model is used to measure consumers' *subjective* perception of system credibility. In terms of interpretation of the learning process, it is somewhat different from the traditional learning framework in which the unobserved knowledge (system biasedness and reliability in our context) is fixed at a certain level.

5.3. Consumer Decisions on Whether to Post and What to Rate

So far, we have presented a general model of how quality disconfirmation is formulated, how consumers' perceived system credibility is updated, and how these two mechanisms are linked to each other. In this section, we discuss how we model consumers' online rating decisions.

Propensity Model. In this paper, we posit that whether a consumer would post a rating after product experience is mainly impacted by the realized quality she has, quality disconfirmation she encounters, and the updated belief in system reliability she has. Given that consumers' perception of system credibility is modelled in a form of a distribution, we use the mean of the distribution to measure her perceived system credibility. Since the belief in system reliability follows a gamma distribution, an individual *i*'s expectation of system reliability at time *t*, Rel_{it}, is given by:

$$Rel_{it} \equiv E\left[\tau_{it}\right] = \frac{a_{it} - 1}{b_{it}}. \tag{14}$$

[15] The choice of a_{i0} is subjective. We estimate the proposed model with a_{i0} fixed at different values (2.5, 5, 10, and 20) and we do not observe noticeable changes for the estimated parameters. Given this result, we report the estimation results with $a_{i0} = 5$ since it gives the best model fit.

Given the belief updating rule provided by (12) and (13), the latent variable Rel_{it} will become larger if the magnitude of disconfirmation is relatively small, meaning that the individual perceives the review system to be more reliable. To investigate how an individual's idiosyncratic belief in system reliability affects her posting behavior, we incorporate the latent construct Rel_{it} into the raters' decision-making process.

We model that a consumer's decision of whether to post a product rating is governed by latent *posting propensity*. Specifically, individual i's posting propensity for product j purchased at time t, $Prop_{ijt}$, is expressed as:

$$\begin{aligned} Prop^{*}_{ijt} &= Prop_{ijt} + \varepsilon_{p,ijt} \\ &= \beta_{i0} + \beta_1 \Delta\bar{Q}_{ijt} + \beta_2 \Delta\bar{Q}^2_{ijt} + \beta_3 Rel_{it} + \beta_4 Q_{ijt} + \beta_5 Q^2_{ijt} + \beta_{6:8} X_{jt} + \varepsilon_{p,ijt}, \end{aligned} \tag{15}$$

where β_{i0} allows for heterogeneity in baseline propensity across individuals and can be interpreted as individual-specific net utility the individual i derives from posting. The covariate ΔQ_{ijt} is the mean of quality disconfirmation given in (5), Rel_{it} is the time-variant belief variable specified in (14), Q_{ijt} is the realized quality given in (2). We also add quadratic terms, $\Delta\bar{Q}^2_{ijt}$ and Q^2_{ijt}, to capture possible nonlinear relationship between $Prop_{ijt}$ and quantities of our interest. To control the price effect and the impact of rating environments on a focal individual's posting decision, we include product price, volume and the statistical variance of posted ratings in X_{jt}.[16] Finally, $\varepsilon_{p,ijt}$ is an idiosyncratic error with a mean of 0. Since the outcomes of posting decision is binary (posting or lurking), we assume $\varepsilon_{p,ijt} \sim N(0,\ 1)$ and the resulting propensity model has a standard probit specification:

$$\Pr(y_{ijt} = 1) = \Pr(Prop^{*}_{ijt} > 0) = \Phi\left(Prop_{ijt}\right), \tag{16}$$

where y_{ijt} is an occasion-specific dummy indivating whether an individual i posts a product rating for j purchased at time t.

The parameters of our main interest are β_1, β_2, and β_3. The estimates of β_1 and β_2 together will inform us the impacted of quality disconfirmation on consumers' posting decision. The estimated parameter β_3 will indicate whether one's perceived system reliability would encourage (if $\beta_3 > 0$) or deter (if $\beta_3 < 0$) herself to express her own opinions about the product. Moreover, it has been shown that online opinions are subject to a polar effect – consumers with extreme opinions tend to be more vocal (Anderson 1998; Dellarocas and Narayan 2006). We will be able to confirm the existence of polar effect if estimated β_4 is negative whereas β_5 is positive.

[16] We observe from our data that a typical consumer either posts a product rating within two weeks after the order has been placed or does not express their opinions at all. The rating environments (such as volume, valence and variance of ratings) does not have noticeable changes during such short period of time. Therefore we use review statistics observed at time t to characterize the rating environment individual i is exposed to.

Evaluation Model. Now, consider that the individual i has decided to post a proudct rating for j. We assume that i's decision of what score to rate is governed by latent *rating evaluation*, which is mainly driven by her realized quality:[17]

$$Eval^*_{ijt} = Eval_{ijt} + \varepsilon_{e,ijt} = Q_{ijt} + \lambda_1 h_{ijt} + \lambda_{2:4} X_{jt} + \varepsilon_{e,ijt}\,, \tag{17}$$

where $\varepsilon_{e,ijt}$ is a zero-mean random shock. While consumers are advised to provide product-related feedback only, we suspect that some consumers may reflect the level of service they receive from the e-commerce site in product ratings. To account for such possibility, we include log-transformed order handling time (measured by days), h_{ijt}, in the formulation of rating evaluation. Since rating evaluation is continuous whereas the ordinal ratings only take integer values (1 to 5 in our case), we assume that the relationship between them follows:

$$z_{ijt} = \begin{cases} 5, & if \;\; \kappa_4 < Eval^*_{ijt} < \kappa_5, \\ 4, & if \;\; \kappa_3 < Eval^*_{ijt} \le \kappa_4, \\ 3, & if \;\; \kappa_2 < Eval^*_{ijt} \le \kappa_3, \\ 2, & if \;\; \kappa_1 < Eval^*_{ijt} \le \kappa_2, \\ 1, & if \;\; \kappa_0 < Eval^*_{ijt} \le \kappa_1, \end{cases} \tag{18}$$

where z_{ijt} denotes the observed rating scores and κ. specifies the utility-ratings translating cutpoints. For identification purpose we set $\kappa_0 = -\infty$, $\kappa_1 = 0$ and $\kappa_5 = \infty$ (Koop et al. 2007), resulting in three cutpoints κ_2, κ_3 and κ_4 to be estimated.

5.4. Interdependence between Two Rating Decisions

So far we have developed two separate models governing raters' decisions of whether to post and what to rate. However, the covariance matrix of the equation system has not yet been clearly specified. Based on the independence assumption among two decision stages,[18] MS assume that the idiosyncratic error ε_p in propensity model and ε_e in evaluation model both follow a standard normal distribution and are independent of each other. However, since the realized quality Q_{ijt} enters two models simultaneously, all parameter estimates could be biased if the interdependence between two decisions is not explicitly specified. As a result, we assume two sets of error terms to follow a bivariate distribution:

$$\begin{pmatrix} \varepsilon_p \\ \varepsilon_e \end{pmatrix} \sim BVN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right). \tag{19}$$

[17] In other words, we assume that the posted ratings are trustful and serve as an unbiased signal of consumers' product evaluation.

[18] They make this simplification mainly because from estimation results they find that correlation between two error terms is small and the parameter estimates are not substantively different.

For identification purpose, we fix the standard deviations of ε_p and ε_e at 1 in order to obtain binary probit and ordered probit specification, respectively. The parameter ρ is the correlation coefficient to be estimated[19]. Given this covariance structure and the translating cutpoints defined in (18), the probability of observing a joint event of $y_{ijt} = 1$ and $z_{ijt} = s$ is given by:

$$\Pr(y_{ijt}=1, z_{ijt}=s) = \begin{cases} \Phi_2(\infty, Prop_{ijt}, \rho) - \Phi_2(\kappa_4 - Eval_{ijt}, Prop_{ijt}, \rho) & s=5, \\ \Phi_2(\kappa_4 - Eval_{ijt}, Prop_{ijt}, \rho) - \Phi_2(\kappa_3 - Eval_{ijt}, Prop_{ijt}, \rho) & s=4, \\ \Phi_2(\kappa_3 - Eval_{ijt}, Prop_{ijt}, \rho) - \Phi_2(\kappa_2 - Eval_{ijt}, Prop_{ijt}, \rho) & s=3, \\ \Phi_2(\kappa_2 - Eval_{ijt}, Prop_{ijt}, \rho) - \Phi_2(\quad - Eval_{ijt}, Prop_{ijt}, \rho) & s=2, \\ \Phi_2(\quad - Eval_{ijt}, Prop_{ijt}, \rho) - \Phi_2(-\infty, Prop_{ijt}, \rho) & s=1, \end{cases} \tag{20}$$

where Φ_2 denotes the standard bivariate normal cumulative distribution function (CDF). Given such specification, our proposed model releases the independence assumption between two observed outcomes (y's and z's). The probability of observing $y_{ijt} = 0$ (i.e. lurking) can be expressed as:[20]

$$\Pr(y_{ijt}=0) = 1 - \Phi\left(Prop_{ijt}\right). \tag{21}$$

Based on (20) and (21), the joint likelihood for observing individual i who makes m purchase occasions and posts n product ratings is given by

$$L_i\left(y_{i\cdot}, z_{i\cdot}\right) = \prod_{t\in(y_{ijt}=0)} \Pr\left(y_{ijt}=0\right) \cdot \prod_{t\in(y_{ijt}=1)} \Pr\left(y_{ijt}=1, z_{ijt}=s\right), \tag{22}$$

|--- ($m - n$) terms ---| |---------- (n) terms ---------|

and the likelihood of observing the entire decision set made by N individuals is $\prod_{i=1}^{N} L_i\left(y_{i\cdot}, z_{i\cdot}\right)$.

6. Estimation

We generalize the second stage of the Heckman selection model to reflect ordinal outcomes (rating scores in our context) and directly estimate two interdependent outcomes using a bivariate normal specification for correlated error terms. For parameters that are subject to certain constraints we apply the following transformation strategy. First, since the correlation coefficient ρ takes value from [–1, 1] only, we estimate the inverse arctangent transformation of it, which maps the support [–1, 1] to real line (Ying et al. 2006). Second, we estimate $log(\bar{b}_0)$ instead of $\bar{b}_0$ because the inverse scale parameter of a gamma distribution takes positive values only. Finally, to ensure the magnitude of three cutoffs to obey a desired order (i.e. $\kappa_2 < \kappa_3 < \kappa_4$), we estimate the log of difference between two adjacent cutoffs. For variables having hyper-dispersion property such as product price p_{ijt} and order handling time h_{ijt}, we take logarithm transformation. We also mean-center all variables using the grand means to reduce the correlation

[19] To estimate ρ, we first compute inverse Mills ratio, $IMR_{ijt} = \varphi(Prop_{ijt}) / \Phi(Prop_{ijt})$ and plug $\rho \cdot IMR_{ijt}$ into (17).

[20] The marginal distribution of one series of bivariate normally distributed variables is simply a normal density.

between the estimated intercepts and slopes and avoid potential multicollinearity in the propensity model (especially for variables which enter propensity model in both linear and quadratic forms).[21]

To estimate the proposed model, we use hierarchical Bayes approach which allows us to identify parameters at the individual level. Given the nature of parameter hierarchy, the parameters in our model can be divided into two groups (Netzer et al. 2008): (1) "random-effect" parameters that vary across individuals (denoted by θ_i); and (2) "fixed" parameters that do not vary across individuals (denoted by ψ). We allow individual-specific parameters governing propensity intercept, realized quality and initial learning parameter to be correlated by assuming:

$$\theta_i \sim \mathrm{MVN}\left(\bar{\theta},\, \Sigma\right), \tag{23}$$

where $\bar{\theta}$ denotes the mean effects that persist across individuals and Σ denotes the variance and covariance matrix of θ.

We use uninformative priors as we do not have much prior knowledge about model parameters. Specifically, we use diffuse multivariate normal for the fixed parameter ψ. There are 18 elements in the vector ψ, including β capturing the coefficients for 8 covariates in the propensity model, λ measuring the effect of 4 covariates in the evaluation model, cutpoint set (κ_2, κ_3, κ_4) mapping continuous rating evaluation and discrete posted scores, and $\bar{\theta}$ governing the mean effect of baseline propensity and realized quality. A diffuse multivariate normal is used for parameters that vary across individuals. Let ξ_i denote individual-specific deviation from $\bar{\theta}$. Following (23), we can directly estimate those deviations using $\xi_i \sim \mathrm{MVN}(0, \Sigma)$.

We develop a Markov chain Monte Carlo (MCMC) procedure to recursively draw parameters from the corresponding full conditional distributions using the following steps:

$$\begin{aligned} &\xi_i \mid Y_i, Z_i, \psi, \rho, \Sigma; \\ &\Sigma \mid \xi_i; \\ &\psi \mid Y, Z, \xi_i, \rho; \\ &\rho \mid Y, Z, \psi, \xi_i; \end{aligned} \tag{24}$$

We adopt random walk Metropolis-Hasting algorithm for steps where the conditional posterior distributions do not have a closed form (step 1, 3 and 4). To improve the efficiency of the sampler, we use a two-stage strategy. In the first stage, we run the MCMC for 50,000 iterations and discard the first 25,000 draws as "burn-in" samples.[22] We calculate posterior means and empirical variance based on the remaining 25,000 draws. In the second stage, we run a separate MCMC using the posterior means

[21] We calculate the variance inflation factor (VIF) for the covariates included in the propensity model and the result shows no evidence of multicollinearity (VIF < 2).

[22] The choice of burn-in period is based on the visual observation on the trace plot of MCMC draws. In fact, 25,000 is a conservative number since the chain appears to converge after initial 5,000 iterations.

calculated in the previous stage as the initial values and the empirical variance as the diagonal elements of covariance matrix of the random-walk proposal distribution.[23] We run the MCMC sampler for 100,000 iterations and record every 10th draw only to mitigate the autocorrelation issue which is an evitable consequence of the MCMC simulation (Hoff 2009). The adaptive chain converges immediately and explores the parameter space efficiently. To test convergence, we perform Gelman and Rubin diagnostic (Gelman and Rubin 1992) by running 3 parallel chains with different initial values and random seeds. The potential scale reduction factors (PSRF) are below 1.02 for all parameters, suggesting that the chains have converged to the target distributions.[24] We combine draws from three parallel chains, resulting in an effective sample size of at least 500 for all parameters.

6.1. Model Fit and Comparison

To demonstrate the fit of our proposed model, we compare our full (dynamic) model with its three nested versions:

Model 1 – Static model (no dynamic learning).

Model 2 – Static model without quality disconfirmation entering propensity model.

Model 3 – Static model without quality disconfirmation and realized quality entering propensity model.

Table 2. Model Comparison

Model	Dynamic Belief Updating	Quality Disconfirmation	Realized Quality	DIC w.r.t. Full Model*
Full	Yes	Yes	Yes	---
Model 1	No	Yes	Yes	349.7
Model 2	No	No	Yes	427.2
Model 3	No	No	No	483.9

* Lower DIC is preferred.

Table 2 reports the deviance information criterion (DIC) proposed by Spiegelhalter et al. (2002) for four alternative models. The full model outperforms its three nested versions, and therefore we report and interpret the estimation results from the full model in the next section. Our full model has a better model fit over three static models because the belief updating mechanism provides a general way to explain consumers' posting decision over time. In addition, the model fit becomes worse and worse if we exclude the effect of quality disconfirmation (Model 2) and the effect of realized quality (Model 3) one by one. This pattern provides evidence that it is desirable to incorporate all three components when modeling consumers' online rating behavior.

[23] The scale parameter of proposal distribution is adaptively chosen such that the acceptance rate is around 23%, as suggested for high-dimension vector (Gelman et al. 2003).
[24] A series of MCMC draws are considered to achieve convergence as long as PSRF < 1.2.

7. Results

Table 3 reports the posterior means of parameters specified in the propensity model. Since we apply a Bayesian approach to estimate the model, we evaluate the significance of parameter estimates based on the highest posterior density (HPD) intervals. A HPD interval describes the information we have about the location of the true parameter after we have observed the data (Hoff 2009). Parameter estimates are considered significant if the corresponding HPD intervals do not contain zero.

Table 3. Estimation Results (Propensity Model)

Model	Notation	Description	Coeff.	Signif.
Propensity	β_{i0}	Mean effect of baseline propensity	–0.831	**
	β_1	Disconfirmation – Linear	–0.011	
	β_2	Disconfirmation – Quadratic	0.025	**
	β_3	Perceived system reliability	–0.508	**
	β_4	Realized quality – Linear	–0.268	**
	β_5	Realized quality – Quadratic	0.065	**
	β_6	Product Price	0.271	**
	β_7	Volume of posted ratings	–0.009	
	β_8	Variance of posted ratings	0.009	*

** (*) Indicates 95% (90%) HPD interval does not contain 0.

The parameters β_1 and β_2 indicate that online shoppers' propensity of posting product ratings is , at least partially, driven by quality disconfirmation. That is, the larger the extent to which the *ex-post* perceived quality deviates from the *ex-ante* expectation, the more likely a consumer is to express her own product opinion. If the average rating she observes more or less reflects her own post-purchase evaluation, she is more prone to "lurk". Such behavior could be attributed to many reasons such as altruism. An altruist may praise the product by submitting a rating higher than the current mean rating if her product evaluation is actually higher than the level represented by the rating signal. In contrast to, if her perceived product quality is lower than what review signal indicates, she may warn peer consumers to be vigilant with the product by posting a below-the-average rating. Moreover, the coefficient is significant and positive for quadratic term (β_2) while insignificant for linear term (β_1), suggesting that the effect of disconfirmation on posting propensity is increasing in the magnitude of disconfirmation. While existing work has documented several motives why consumers engage in WOM from a normative point of view (Hennig-Thurau et al. 2004; Sundaram et al. 1998), this research provides a positive validation and contribute to literature with an economic-based explanation: the disconfirmation effect identified in this paper serves as one of the underlying forces driving consumers to voluntarily engage in online WOM.

A unique feature of our proposed model is that it models how online raters' posting propensity changes as their perceived system reliability evolves over time. The estimation results show that consumers are less active in participating in WOM activities ($\beta_3 < 0$) as they perceived the review system to be more reliable. Alternatively, we can say that consumers tend to be more vocal when they do not trust rating signals, relatively to the case where they perceived the system to be highly reliable. It should be clear that occasion-specific disconfirmation has a different economic meaning from the perception of system reliability. The former captures the short-term impact of a disconfirmation shock on the corresponding review-posting opportunity, whereas the latter measures how an individual perceives the system based on all of her past disconfirmation experiences and can be considered to have a long-term and accumulative effect on one's posting behavior.

As for the effect of the realized quality on posting propensity, the estimated coefficient is negative for the linear term (β_4) but positive for the quadratic term (β_5). Combined, these two parameter estimates indicate that consumers are more likely to share opinions when they perceive product quality to be either very high or very low. This finding echoes the polar effect observed by Anderson (1998) in an offline setting and by several researchers in an online setting (Dellarocas et al. 2010; Dellarocas and Narayan 2006; Moe and Schweidel 2012). Furthermore, we also find that consumers are more interested in rating products with higher prices ($\beta_6 > 0$). The rating environment, characterized by the volume and variance of posted ratings, has dissimilar impacts on a focal consumer's posting decision. In particular, online raters' posting decision is subject to a crowding-out effect, meaning that a focal individual tends to lurk if there has been a big crowd sharing their opinions ($\beta_7 < 0$). Such finding is analogous to a political phenomenon where voters tend to abstain if public polls have declared clear winners (Sudman 1986). On the other hand, the dissension of posted opinions would encourage the focal individual to share her own opinion ($\beta_8 > 0$).

The estimation results of other parameters specified in the evaluation model, utility-score translating rule and system credibility updating model are reported in Table 4. As expected, the order handing time has a negative impact on rating evaluation ($\lambda_1 < 0$). This shows that online shoppers commonly reflect the service level they receive from the e-commerce site in the product ratings. The coefficient for product price is positive but insignificant, meaning that price in general is a weak proxy for quality. Unlike in the propensity model, the dissension of public opinions and latent rating evaluation is negatively correlated. For the ease of interpretation, we report the transformed cutoff points of the utility-score translating rule. The posterior mean of our estimate of $\log(b_i)$ is significant, suggesting that the updating mechanism well captures the dynamic evolution of how consumers perceive the reliability of the review system. Finally, the interdependence between two rating stages is positive yet insignificant.

Table 4. Estimation Results of Other Mean-Effect Parameters

Model	Notation	Description	Mean	Signif.
Evaluation	λ_{i0}	Mean effect of realized quality	1.997	**
	λ_1	Order handling days	–0.104	**
	λ_2	Product Price	0.033	
	λ_3	Volume of the rating environment	–0.015	
	λ_4	Variance of the rating environment	–0.033	*
Cutoffs	κ_1	Cutpoints for s=1 and 2 (fixed at 0)	0.000	---
	κ_2	Cutpoints for s=2 and 3	0.486	**
	κ_3	Cutpoints for s=3 and 4	1.406	**
	κ_4	Cutpoints for s=4 and 5	2.459	**
Credibility Updating	$\log(b_{i0})$	Mean effect of initial inverse scale parameter (log-transformed)	2.189	**
	a_{i0}	Initial shape parameter (fixed at 5)	5.000	---
Correlation	ρ	Interdependency between two rating stages	0.082	

** (*) Indicates 95% (90%) HPD interval does not contain 0.

Table 5 reports the mean and standard deviation among individual-specific parameters. The large standard deviations for β_{i0} and λ_{i0} suggest that online reviewers are substantially heterogeneous in baseline posting propensity and rating evaluation (also see Figure 5). To explore the interdependence between baseline latent parameters at the individual level, we provide the pair-wise correlation matrix[25] among them in Table 6. The positive correlation between β_{i0} and λ_{i0} suggests that after controlling for other factors, online consumers who have higher baseline posting propensity appear to be more lenient in terms of giving numeric ratings. This finding is opposite to Moe and Schweidel (2012) who observe that active raters are more negative. Finally, our estimate results indicate less heterogeneity across individuals in their initial perception of system reliability, as suggested by a small standard deviation for b_{i0}.

Table 5. Individual-specific Parameter Estimates

Model	Notation	Mean among Individuals	Std. dev. among individuals
Baseline propensity	β_{i0}	–0.831	0.598
Baseline evaluation	λ_{i0}	1.997	0.520
Initial learning status	b_{i0}	8.931	0.208

[25] Alternatively, we can evaluate the posterior estimates of covariance.

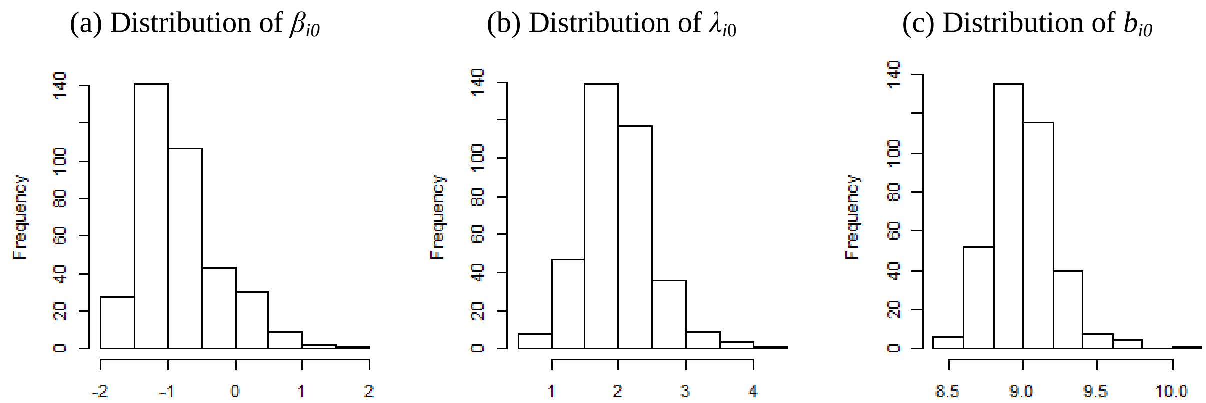


Figure 5. Distribution of Individual-specific Parameters

Table 5. Individual-specific Parameter Estimates

	β_{i0}	λ_{i0}	b_{i0}
β_{i0}	1.000	0.142	0.387
λ_{i0}		1.000	0.876
b_{i0}			1.000

7.1. Extension and Robustness Checks

In this section, we relax some of our model assumptions by incorporating other factors that might affect online raters posting behavior into our proposed model. The estimation results are consistent before and after we consider the following two robustness checks.

7.1.1. The Effect of Quality Disconfirmation on Product Evaluation

In our proposed model, we assume that quality disconfirmation enters the propensity model directly and does not have any impact on consumers' realized quality. That is, consumers do not factor their prior expectation about quality into product evaluation. In fact, marketing literature finds that in offline settings, a consumer overall satisfaction is affected by quality disconfirmation (Anderson and Sullivan 1993) and such relationship is empirically shown to be positive (Rust et al. 1999). Based on this theory, we now assume a consumer's rating evaluation is mainly driven by her overall product satisfaction S, which is a linear combination of her realized quality and quality disconfirmation:

$$Eval_{ijt} = S_{ijt} + \lambda_1 h_{ijt} + \lambda_{2:4} X_{jt}, \tag{25}$$

where $S_{ijt} = Q_{ijt} + \gamma \cdot \Delta Q_{ijt}$ and γ is a new parameter to be estimated. With this specification, the posting propensity is now written as:

$$Prop_{ijt} = \alpha_i + \beta_1 \Delta\bar{Q}_{ijt} + \beta_2 \Delta\bar{Q}_{ijt}^2 + \beta_3 Cred_{it} + \beta_4 Q_{ijt} + \beta_5 Q_{ijt}^2 + \beta_{6:8} X_{jt}. \tag{26}$$

It can be shown that the mean of quality disconfirmation becomes:

$$\Delta\bar{Q} = (\tilde{R}_j + \lambda_{i0}) - (\bar{R}_{jt} - \delta_{it}) / (1+\gamma). \tag{27}$$

Comparing (27) with (5), we can see that the inclusion of γ will bring the mean value of disconfirmation signal towards zero, provided $\gamma > 0$.

The estimation results (including parameter estimates and DIC) of the new model only have marginal deviation from those reported in Table 3, suggesting that the main results of this paper is robust even if we consider the impact of disconfirmation on overall product satisfaction. The posterior mean of γ is 0.117 and is significant at 95% level. This result indicates that online rates tend to factor in their expected product quality when leaving product ratings.

7.1.2. The Interaction between Disconfirmation and Rating Environments

We have examined how a focal consumer's posting decision is influenced by disconfirmation she encounters as well as rating environments she is exposed to. An interesting question to ask is: Is the disconfirmation effect itself also moderated by opinions expressed by the crowd? To answer this, we allow the effect of disconfirmation on posting propensity to interact with variables that characterize the rating environment (i.e. volume and variance of posted ratings) and add these two interaction terms into our propensity model. The parameters of interest are reported in table 7.[26]

Table 7. Effects of Rating Environments on Posting Behavior

Effect	Coeff.	Signif.
Volume on propensity	–0.008	
Variance on propensity	**0.023**	**
Volume on Disconfirmation Effect	0.026	
Variance on Disconfirmation Effect	**–0.008**	*

** (*) Indicates 95% (90%) HPD interval does not contain 0.

Consistent with the results shown in Table 3, we still find the evidence that the variance of posted ratings induces posting behavior from our extended model. Interestingly, the coefficient for "Disconfirmation × Variance" interaction term is negative and significant. This indicates that the effect of disconfirmation is weakened by the level of dissension among posted opinions, perhaps because a consumer anticipates the review information to be noisy and hence become less sensitive to the disconfirmation effect. Moreover, the coefficient of "Disconfirmation × Volume" interaction has a positive sign. This also makes sense. The higher the volume of posted ratings, the more trustworthy consumers will perceive the review signal, and therefore the more sensitive they will be to the disconfirmation effect.

[26] Since the parameter estimates do not have a substantial change, here we report coefficients for variables that are related to rating environments only.

7.1.3. The Use of the Long-term Average Rating as a Proxy for Product Quality

We use the long-term average rating to proxy true product quality in our realized quality equation. Although we have justified that such proxy is reasonable (in Footnote 10), one may argue that the long-term average rating could still be a biased signal to some extent. Hu et al. (2006) argue that the average product rating would over-report (under-report) the true quality if consumers as a whole are more inclined to brag (moan) about the product when they are highly satisfied (disgruntled). To address this issue, we add a product-level random effect in the realized quality function. As a result, (2) can be rewritten as:

$$Q_{ij} = (\tilde{R}_j + \varpi_j) + \lambda_{i0}. \tag{28}$$

Parameter ϖ_j models the randomness that some products are overrated whereas some are underrated such that $\varpi_j \sim N(0, \sigma_{\varpi}^2)$. If the value of the parameter to be estimated σ_{ϖ} is small, we can conclude that the long-term average ratings more or less reflect true product quality.

The estimation results obtained from this new specification do not have noticeable difference from those obtained from our proposed model. The posterior mean (standard deviation) for ϖ_j is 0.011 (6×10^{-5}), providing supportive evidence that the long-term average rating well captures the true product quality. What makes our finding different from the extant work is that we explicitly consider consumers' posting behavior is driven by quality disconfirmation (in addition to the realized quality) and such disconfirmation "ensures" that the average rating will gradually converge to the true quality.[27] On the contrary, the conclusion by Hu et al. (2006) is contingent on an analytical assumption that consumers' decisions of whether to post are solely determined by perceived quality.

7.2. Simulations and Analyses on Consumers' Rating Behavior

In this subsection we run two sets of simulations to further highlight the significance of our findings and better understand the online reviewers' rating behavior.

7.2.1. Recovering unobserved rating scores

Since our model explicitly considers online raters' decisions of whether to rate and what score to rate for the product, we are able to compute the latent posting propensity and evaluation for every purchase occasion. Utilizing the estimated cutoff points, we can further recover the unobserved rating scores for purchase occasions that do not lead to review submissions. The simulation procedure for recovering unobserved ratings is summarized as follows:

1. We compute the latent posting propensity per (15) and evaluation per (17) for all purchase occasions (including those lurking sessions) based on the posterior estimates.

[27] We have a further discussion on the convergence of average ratings in Section 7.2.2.

2. Given the computed posting propensity, evaluation and estimated cutpoints, we simulate raters' decisions of whether to post per and what score to rate per (16) and (18).
3. We repeat Step 2 for 1,000 iterations and compute the average for the quantities of our interests across iterations.
4. To make sure the number of iterations is sufficient, we repeat Steps 2-4 in three parallel processes with different random seeds. We compare simulated results and do not find any inconsistency across processes.

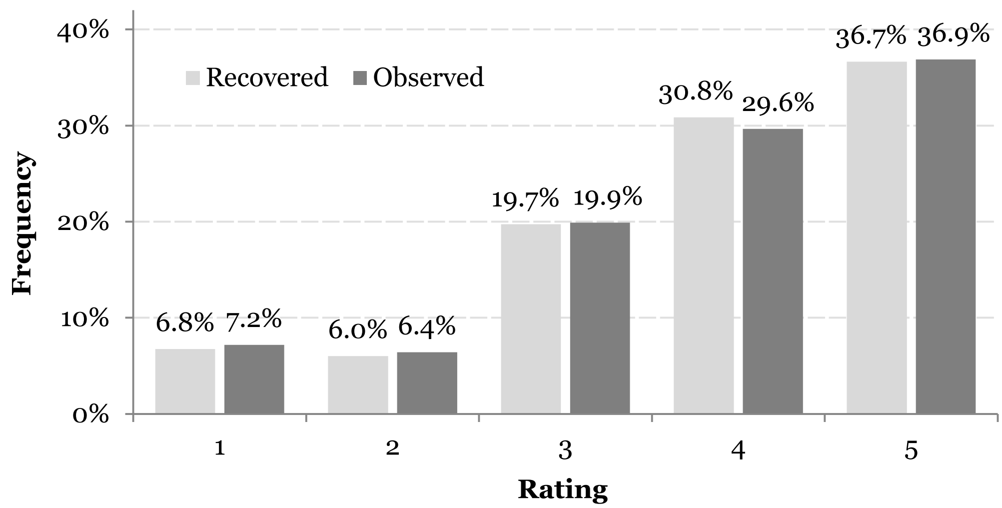


Figure 7. Simulated vs. Observed Ratings

We begin our analysis by assessing the performance of our model on describing online reviewers' rating behavior. Figure 7 plots the frequency of simulated scores for occasions with review entries (presented by light grey bars) against the frequency of ratings observed in the data set (presented by dark grey bars). Our simulated results are very close to observed outcomes, indicating that the proposed model well captures the mechanism governing consumers' rating decisions.[28]

To further understand the relationship between posting propensity and perceived product quality, we recover the unobserved scores for missing rating points. For interpretation purpose, we define *posting percentage* of a score *s* as $\frac{\textit{frequency of a score s being posted}}{\textit{frequency of a score s being evaluated}}$ and plot the posting percentage for each of 5 scores in Figure 8. We find that the relationship between posting percentage and discretized product evaluation is best characterized by a left-skewed U shape. On average negative experiences (rated as 1 or 2) have a higher chance to be reported (~10.69%), relative to 8.59% for neutral and roughly 7.95%

[28] We also run the same analysis for models without inclusion of quality disconfirmation (defined in Section 6.1). The results show that the proposed model outperforms other alternatives in terms of explaining rating behavior as well.

for positive evaluations. Such systematic difference is primarily caused by the stronger effect of negative disconfirmation. It is worth noting that the prior research observing a conventional U-shaped relationship is grounded on overall consumer satisfaction, whereas in this research we further decompose satisfaction into product evaluation and quality disconfirmation (Anderson and Sullivan 1993) and examine the impacts of both components on eWOM contribution. Such modeling advantage allows us to obtain a sharper insight into online raters' behavior and further confirm the U-shaped or J-shaped distribution of ratings commonly observed in online product review systems.

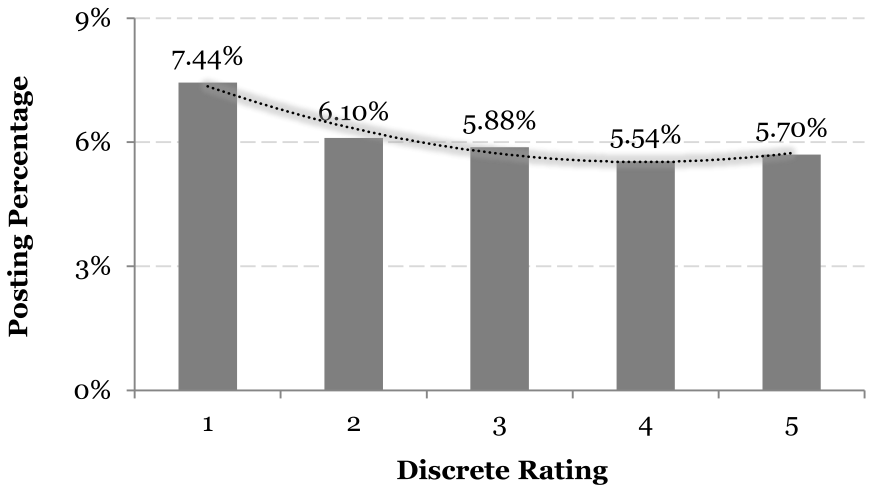


Figure 8. Posting Percentage Breakdown by Score

7.2.2. Convergence of product ratings and the effect of review manipulation

To highlight the effect of disconfirmation and the evolution of online product ratings, we perform our second simulation by following the procedure below.

1. We sample 3,000 individuals from the posterior estimates of population-level parameters (Σ).
2. For each individual we simulate her decision of whether to post; for individuals who decide to rate, we simulate her decision of what score to rate. If a new rating is posted, we calculate the up-to-date mean ratings as well as other rating environment variables that will be used as control variables in the subsequent rating occasion.
3. We repeat Step 2 until 150 ratings are posted.
4. We repeat Steps 1-3 for multiple iterations.

Figure 7 plots different patterns of the evolution of online product ratings identified in our simulation results. It is evident that in the long run the mean rating will converge to its true quality (a 3.5 star in our simulation setting). If earlier buyers over-evaluate the product, perhaps due to the self-selection biases (Li and Hitt 2008), the inflated rating would urge disappointed consumers to express their negative opinions in the later period (see Figure 9a). While such declining trend of online ratings have been attributed to different effects such as self-selection biases (Li and Hitt 2008) and environmental factors (Moe and

Schweidel 2012), the effect of quality disconfirmation identified in this research can also provide an alternative explanation. Perhaps more importantly, the disconfirmation effect also can be used to explain other commonly observed evolution patterns. For example, if the product is initially underrated, the subsequent buyers who derive positive disconfirmation would praise the product by sharing their pleasant experiences (as shown by Figure 9b). Interestingly, we also observe an undershooting property where the average rating first exhibits a steep declining pattern and then bounces back to a steady state (see Figure 9d). We believe that the quality disconfirmation is one of the most important forces driving consumers' posting behavior and such effect is also validated by different evolution paths of online ratings commonly observed across contexts and platforms.

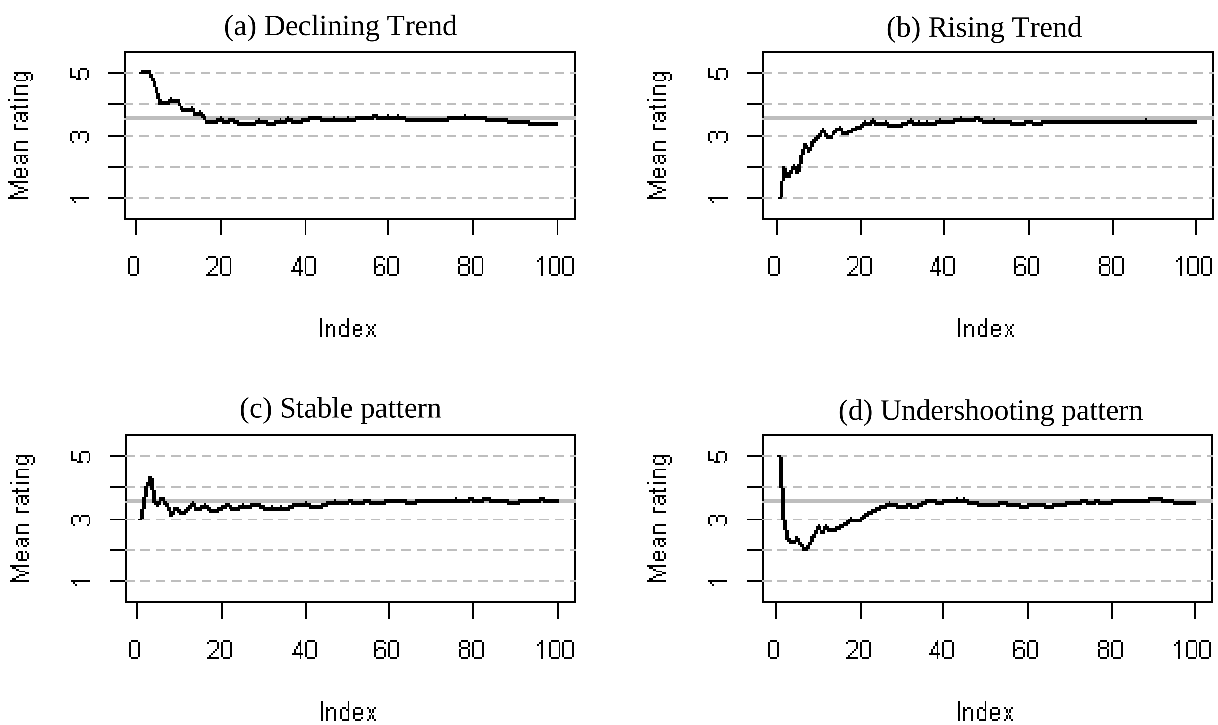


Figure 9. Various evolutionary patterns of online product ratings

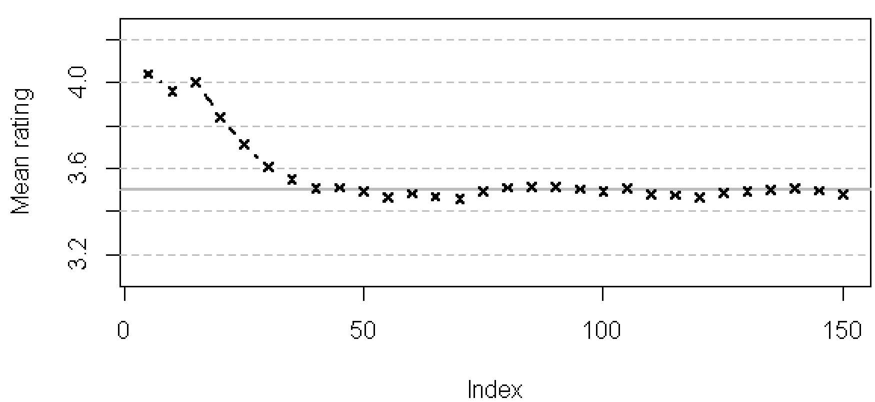


Figure 10. Evolution of average rating in the presence of inflated ratings in the introductory period

Our simulation can also be used to understand the effect of strategic manipulation of online product ratings. Similar to previous setting, we consider a scenario where the seller is able to strategically manipulate customer reviews such that the average rating will be as high as 4 at the 15-th entry (an inflation of a half star). We repeat Steps 1-4 listed above for 10 iterations and average the posted ratings cross iterations. As Figure 10 shows, while the mean rating can be inflated in the early stage through manipulation, it quickly drops and converges to the true quality (3.5). This result suggests that although fake online reviews can elevate prospective buyers' prior expectation, such effect would not last long as more and more disgruntled consumer diffuse negative WOM in order for anxiety reduction or vengeance purpose. On the other hand, we can also expect that the damage caused by the malicious review manipulation by competing firms should be alleviated by subsequent positive ratings as well.

8. Conclusion

This paper examines the online reviewers' rating behavior and attempts to explain several phenomena commonly observed across online rating systems. The early research on online product reviews has studied how user-generated ratings can be related to market performance whereas the recent work focuses on the impact of existing ratings on subsequent ones from a social dynamics standpoint (Lee et al. 2013; Moe and Schweidel 2012; Shen et al. 2013). Our work contributes to the latter literature stream by proposing a novel framework in which a focal consumer's posting decisions has been influenced by others' opinions before the product is purchased and consumed. In addition, we model how consumers' perception of the system credibility evolves over time. By integrating these two features with other factors such as environmental and price effects, we are able to examine what drives consumers to voluntarily contribute online product ratings from a more comprehensive viewpoint.

Using a rich data set containing complete purchase and review history at the individual level, we empirically show that the rating behavior is driven by the degree to which the realized quality deviates from the expected quality obtained from the same product. When interacting with the rating environments, the intensity of such effect is decreasing in the variance of posted ratings while increasing in the volume of submitted reviews. We also find that online raters tend to be more vocal when they perceive the review system to be less credible.

The main finding of this paper echoes prior research in some aspects but also provides different insights in others. On one hand, the disconfirmation effect can serve as the underlying driver of why people engaging in WOM such as concerns for others, anxiety reduction, vengeance, etc (Anderson 1998; Hennig-Thurau et al. 2004). The impact of quality disconfirmation on posting behavior can also (at least partially) explain 1) the commonly observed U-shaped distribution of online product ratings; and 2) the declining trend of average ratings at product level. On the other hand, based on our finding we believe that the long-term average rating can represent the true product quality. Through simulations, we demonstrate that consumers' perceived disconfirmation will "ensure" that the average rating converges to the true value in the long run. Our proposition to some extent disagrees with Hu et al. (2006) who argue that the mean score may provide misleading recommendation, which is contingent on an analytical assumption that consumers' posting decisions are solely triggered by their perceived quality.

Our empirical results shed light on the economic value of online product ratings in the following aspects. Manufactures or service providers should be aware that online reviewers are more prone to provide feedback when expectations do not match their own perceived quality. While manipulating product reviews by inflating numeric ratings can temporally boost the sales revenue, fake reviews could turn to be detrimental in the long run as disappointed customers engage in negative eWOM.[29] Furthermore, online reviewers' perceptions of system reliability have a negative impact on their propensity to contribute reviews. Existing literature also shows that rating environment with smaller volume or lower valence discourages posting incidence (Moe and Schweidel 2012). To keep the online rating environment healthy and prosperous, policy makers and marketers who are in charge of online ratings campaigns should design various incentives for different consumers accordingly.

This study has some limitations and can be improved in several directions. First, our data has limited information which prevents us from examining online rating behavior in a more detailed way. For example, although we have utilized the aggregate review information, such as valence and variance, in formulating quality variables, we do not consider the possibility that some consumers may value positive ratings and negative ones differently. If we were able to observe the distribution of posted ratings at the time of purchase, we might be able to discover more interesting findings on the way consumers interpret review signal in the pre-purchase stage. Second, we think we have not yet fully utilized our valuable data. A promising direct for future research is to apply text mining techniques to extract the sentiments stored in the textual data and incorporate them into our econometric model. The synergy created by the integration of different research methods may allow us to provide a shaper insight into consumers' online rating behavior.

Reference

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

Anderson, E.W., and Sullivan, M.W. 1993. "The Antecedents and Consequences of Customer Satisfaction for Firms," *Marketing Science* (12:2), March 20, 1993, pp. 125-143.

Chaiken, S. 1980. "Heuristic Versus Systematic Information Processing and the Use of Source Versus Message Cues in Persuasion," *Journal of personality and social psychology* (39:5), p. 752.

[29] In our empirical model, we can only show that inflated ratings would drive disappointed buyers to leave negative numeric ratings. However, we cannot gauge the harm those negative textual reviews would do to the firm's reputation.

DeGroot, M.H. 1970. *Optimal Statistical Decisions*. Wiley-Interscience.

Dellarocas, C. 2006. "Strategic Manipulation of Internet Opinion Forums: Implications for Consumers and Firms," *Management Science* (52:10), October 1, 2006, pp. 1577-1593.

Dellarocas, C., Fan, M., and Wood, C. 2004. "Self-Interest, Reciprocity, and Participation in Online Reputation Systems,").

Dellarocas, C., Gao, G., and Narayan, R. 2010. "Are Consumers More Likely to Contribute Online Reviews for Hit or Niche Products?," *Journal of Management Information Systems* (27:2), pp. 127-158.

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.

Deloitte. 2007. "Most Consumers Read and Reply on Online Reviews; Companies Must Adjust," USA.

Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. 2003. *Bayesian Data Analysis*. CRC press.

Gelman, A., and Rubin, D.B. 1992. "Inference from Iterative Simulation Using Multiple Sequences," *Statistical science*), pp. 457-472.

Ghose, A., Ipeirotis, P.G., and Li, B. 2012. "Designing Ranking Systems for Hotels on Travel Search Engines by Mining User-Generated and Crowdsourced Content," *Marketing Science* (31:3), pp. 493-520.

Heckman, J.J. 1979. "Sample Selection Bias as a Specification Error," *Econometrica: Journal of the econometric society*), pp. 153-161.

Hennig-Thurau, T., Gwinner, K.P., Walsh, G., and Gremler, D.D. 2004. "Electronic Word-of-Mouth Via Consumer-Opinion Platforms: What Motivates Consumers to Articulate Themselves on the Internet?," *Journal of interactive marketing* (18:1), pp. 38-52.

Hoff, P.D. 2009. *A First Course in Bayesian Statistical Methods*. Springer.

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 conference on Electronic commerce*. Ann Arbor, Michigan, USA: ACM, pp. 324-330.

Koop, G., Poirier, D.J., and Tobias, L. 2007. "Bayesian Econometric Methods (Econometric Exercises)." Cambridge University Press.[Companion to Koop (2003) with lots of hands-on exercises and MATLAB examples].

Lee, Y.J., Tan, Y., and Hosanagar, K. 2013. "Do Ifollow My Friends or the Crowd? Information Cascades in Online Movie Rating." Working paper: Working paper.

Leggatt, H. 2011. "24% of Consumers Turned Off after Two Negative Online Reviews." Retrieved April 13, 2011, from http://www.bizreport.com/2011/04/27-of-consumers-turned-off-by-just-two-negative-online-revie.html

Li, X., and Hitt, L.M. 2008. "Self-Selection and Information Role of Online Product Reviews," *Information Systems Research* (19:4), December 1, 2008, pp. 456-474.

McGlohon, M., Glance, N., and Reiter, Z. 2010. "Star Quality: Aggregating Reviews to Rank Products and Merchants," *International Conference on Weblogs and Social Media*.

Moe, W.W., and Schweidel, D.A. 2012. "Online Product Opinions: Incidence, Evaluation, and Evolution," *Marketing Science* (31:3), May 1, 2012, pp. 372-386.

Netzer, O., Lattin, J.M., and Srinivasan, V. 2008. "A Hidden Markov Model of Customer Relationship Dynamics," *Marketing Science* (27:2), March 1, 2008, pp. 185-204.

Nielson. 2007. "Word-of-Mouth the Most Powerful Selling Tool." Retrieved 2013/04/29, from http://nz.nielsen.com/news/Advertising_Oct07.shtml

PowerReviews. 2010. from http://www.powerreviews.com/products/social-content-engine/customer-reviews

Rust, R.T., Inman, J.J., Jia, J., and Zahorik, A. 1999. "What You Don't Know About Customer-Perceived Quality: The Role of Customer Expectation Distributions," *Marketing Science* (18:1), pp. 77-92.

Schlosser, A.E. 2005. "Posting Versus Lurking: Communicating in a Multiple Audience Context," *Journal of Consumer Research* (32:2), pp. 260-265.

Shen, W., Hu, Y.J., and Rees, J. 2013. "Competing for Attention: An Empirical Study of Online Reviewers' Strategic," *Working paper*).

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., and Van Der Linde, A. 2002. "Bayesian Measures of Model Complexity and Fit," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* (64:4), pp. 583-639.

Sudman, S. 1986. "Do Exit Polls Influence Voting Behavior?," *The Public Opinion Quarterly* (50:3), pp. 331-339.

Sundaram, D.S., Mitra, K., and Webster, C. 1998. "Word-of-Mouth Communications: A Motivational Analysis," *Advances in consumer research* (25:1), pp. 527-531.

Ying, Y., Feinberg, F., and Wedel, M. 2006. "Leveraging Missing Ratings to Improve Online Recommendation Systems," *Journal of Marketing Research* (43:3), pp. 355-365.