Disconfirmation Effect on Online Opinion Expression

Yi-Chun "Chad" Ho • Yong Tan

University of Washington, Seattle, WA 98105

Junjie Wu

Beijing University of Aeronautics and Astronautics, Beijing, China

Abstract

This paper attempts to examine what drives consumers to express their opinions regarding product experience on an e-commerce platform. Using a panel data consisting of complete purchasing and rating history at individual level, we find that the discrepancy between a consumer's expectation and perceived product quality has a significant impact on her propensity to submit a rating, perhaps due to one's intrinsic responsibility to *correct the biased ratings*. Consumers' attention to voice for products exhibits a crowding-out effect, meaning that the responsibility of contributing feedback is dampened if there has been a big crowd expressing their opinions.

Keywords: online product review, rater bias, user-generated content, consumer behavior

[The bulk of the work was done by a student]

1. Introduction

It has been recognized that online product reviews have a substantial impact on consumers' *pre-consumption* decisions such as what products to buy and how the products can fit consumers' need. According to a survey (Nielson 2007), over 75% of the respondents agree that recommendations from peer consumers are the most credible form of advertising. Another survey shows that 43.7% of consumer electronic purchases are affected by online word-of-mouth (BIGResearch 2009). Recently, researchers and practitioners have noticed that this particular form of user-generated content may influence consumers' activities in the *post-consumption* stage. Based on product experience shared by peer consumers, buyers now can formulate a more accurate expectation of the product, leading to higher satisfaction and lower merchandise return rates (PowerReviews). As a result, market efficiency increases due to better product transparency, mainly enhanced by the online product reviews.

A typical online product review system is an information system which facilities information exchange among its users. Based on how they interact with the system, users can be categorized into two groups. The first group of users is information receiver who gathers stored information from various formats, such as overall numeric ratings, textual feedback, etc., and take actions based on the information retrieved from the system. After experiencing the product, an information receiver gets a chance to express her opinions and become an information contributor, the other role a consumer can play while interacting with the system. While how online reviewed can be *used* have been studied in a variety of aspects, lesser attention has been paid to an even more fundamental questions like how online ratings are *generated*. While review postings are commonly assumed to occur completely at random in most research of online ratings, we postulate that decisions on "whether to rate" are subject to a selection mechanism.

Using a panel data consisting of complete transaction and rating history at individual level, we attempt to examine what drives consumers to provide product ratings on an online shopping platform. Our estimate results suggest that the discrepancy between a consumer's expectation and true realization of the product quality significantly influences her propensity to submit a rating. We also find that such tendency is dampened if the volume of submitted ratings available is large. The richness of the data set allows us to examine the evolution of each individual rater's decisions over time instead of looking at the rating sequence observed at product level. The rest of this paper is organized as follows. The next section briefly reviews extant online ratings literature. We then discuss our conceptual model, empirical model, data, followed by estimate results and discussion. Concluding remarks and future research directions are provided in the last section.

2. Literature Review

Driven by the economic value of user-generated content, researchers in IS and marketing have developed a large literature in the "usefulness" of online product ratings. The relationship between different forms of world-of-mouth and product performance has been tested in several contexts such as books (Chevalier and Mayzlin 2006), movies (Dellarocas et al. 2007; Liu 2006), bath and beauty products (Moe and Trusov 2011), etc. A considerable amount of effort has also been spent in order to better understand which aggregate review measures can better predict future sales (e.g., Chintagunta et al. 2010; Clemons et al. 2006; Dellarocas et al. 2007; Duan et al. 2008). Along this line Zhao et al. (2013) model and study consumer learning with online product reviews. Other than those focusing on individual's behaviors, researches conducted from a firm's strategic perspectives propose that firms should adapt their marketing effort and pricing strategy based on consumer reviews (Kuksov and Xie 2010; Sun 2012).

Most of, if not all, works focusing on consumer-generated reviews is based on an implicit assumption: rating is an “unbiased” indicator of the perceived product quality. Seminal works conducted by Dellarocas and Narayan (2006) and Li and Hitt (2008) open a new stream of literature on online rating biases. Dellarocas and Narayan (2006) identify the existence of “*positivity bias*”, based on an observation that online ratings are positively skewed. Li and Hitt (2008) find the “*self-selection bias*” where early buyers of a newly-released book tend to be positively biased in satisfaction, leading to a misrepresentative issue in the early stage. Li and Hitt (2010) further show that unidimentional ratings can be distorted by price effects. Treating missing rating data non-ignorable and informational, Ying et al. (2006) develop a latent model which substantially improves the accuracy of current product recommendation systems. Moe and Schweidel (2012) extend Ying et al’s framework and find that an individual would adjust her own rating decision according to the opinions from the crowd. However, there is still a lack of comprehensive research explaining why online ratings are missed most of the time.

3. Research Framework

In contrast to a common assumption that review incidences occur haphazardly, we posit that the raters’ decision on whether to rate is governed by a selection mechanism. In what follows, we introduce our conceptual model and illustrate how model components affect consumers’ rating behavior.

It has been established in the theory of consumer purchase cycle that consumers first formulate an expected utility for the product before purchase based on information gathered from all kinds of sources, such as product specification, word-of-mouth, advertising, price, etc. This *ex ante* expectation is often referred as pre-consumption utility. In a setting of e-commerce, product ratings provided by the platform has been shown to be the most influential piece of information in determining pre-consumption utility. While the utility formulation is subjective to some extent, it should be clear that a low rating score would have a negative impact on consumers' expectation in mind. After purchasing and experiencing the product, the consumer forms an overall utility based on her own product evaluation and the price effect. In line with the existing literature we call such satisfaction post-consumption utility. The utility disconfirmation, defined as the discrepancy between the pre- and post-consumption utility, realizes right upon the consumer experiences the product.

In this research we postulate that the magnitude of utility disconfirmation will impact consumers' tendency to rate both directly and indirectly. As a review reader, a consumer may want to see that the opinions from the crowd (measured by the averaged ratings) can closely reflect the true quality she would perceive. Upon the post-consumption utility is realized, a consumer has the choice to become a review poster as well by expressing her own opinions. A potential rater may want to "correct" the observed mean ratings, either upwards or downwards, if that averaged rating fails to reflect their perceived quality. Such intention to contribute may be affected by the extent of one's own utility disconfirmation. Furthermore, from a review reader's perspective, consumers' belief in the credibility, or trustworthiness, of the rating system is subject to change over time. If other reviewers' ratings consistently meet one's own experience, she would feel the rating system is trustworthy and therefore her belief in system credibility would accrue over time. On the other hand, if the opinions from the crowd consistently deviate from her own product evaluation, she may find the information contained in the system to be unreliable. As a result, we are also interested in how utility disconfirmation impact consumers' rating behaviors through a dynamic learning process.

To examine online raters' dynamic behaviors, we develop a selection model where decisions on "whether to rate" and "what to rate" are governed by two separate yet correlated mechanisms. We also specify a latent construct to measure the credibility of the system and incorporate it into the proposed selection model.

4. Data

The data for this study is provided by a major online merchant in China who sells a variety of products national wide. This data set contains complete purchase history and rating entries made by 1,000 individuals who are sampled either they make more than 10 purchases, or posts more than 10 ratings, or both. The transactional record set consists of order-specific information such as the product name, price, handling time, order submission date, the status of ratings environment on the date the order is submitted, etc. The rating data set records 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 excellent. Since the data contains complete activity log for all individuals, it spans across a long period of time from January 2008 to October 2011. Participation in product review is completely self-driven. During this period of time, the merchant has not gone through any policy or system design change that could influence consumers' rating behavior.

A major distinction between our data set and those used in prior works is that it allows us to trace the sequence of each individual's purchase and review activity based on the time stamp of each occasion. The richness of this data set enables us to calibrate online rating behavior from a dynamic perspective. This individual-level data also distinguishes our work from others that commonly use review data at product level. To identify the raters' learning dynamics and to make the estimation manageable, we retain users who give three reviews or more. This censored data set contains 142 individuals with 16,841 transactional records and 1,778 review entries. We will use this dataset for model estimation. In June 2013, we also collected a supplementary data set containing the mean ratings for all products appearing in our primary dataset.

Before beginning the model estimation, we briefly present some interesting observations and address the potential issues identified in the data. Figure 1 plots the frequency of 5-scale rating scores submitted by all users in our data set. This positively-skewed distribution is consistent with the pattern commonly found on the online rating platforms (Dellarocas and Narayan 2006; McGlohon et al. 2010). This J-shaped distribution implies that consumers are more prone to express opinions when they are either very satisfied (represented by a high score of 4/5) or very dissatisfied (represented by a low score of 1/2), identified by Anderson (1998). Some variables, such as price and handling time, exhibit a long-tail property. To deal with these over-dispersion issues, we take logarithm transformation on them.

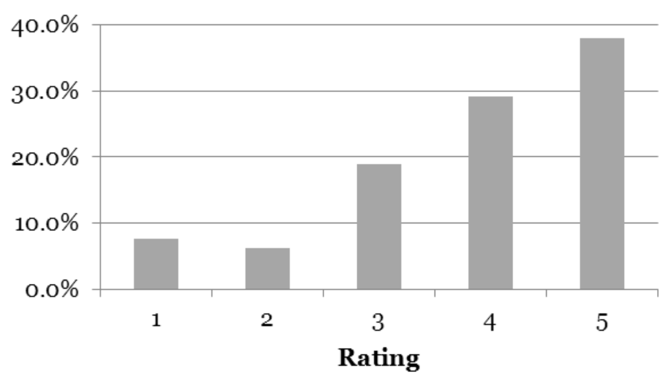

Figure 1. Frequency of Observed Product Ratings

5. Model Development

Following the conceptual framework introduced earlier, we develop a selection model which capture consumers' rating decisions, which are mainly driven by the pre-consumption expectation and post-consumption evaluation of the product. To model how the trustworthiness of the rating system evolves over time, we also develop a latent construct whose value is timevariant, depending on the realization of utility disconfirmation each consumer perceive $\tilde{R}_j$ s. We describe the model in the following order: (1) modeling utility formulation, (2) modeling updating of the credibility of the rating system, (3) modeling consumer decisions on whether to rate and what to rate, and (4) modeling interdependence between two rating decisions.

5.1 Modeling Utility Formulation

Pre-consumption utility. During the research phase of the consumer purchase cycle, consumers gather information for the product from different kinds of sources. In our research context, while a consumer is not able to exactly know how much she will enjoy the product before actually experiencing it, she can formulate a pre-consumption expectation based on the experience shared by peer consumers. Suppose now the expected utility an individual i formulates for product j at time t, $E[U_{ijt}]$, follows the following distribution:

$$E[U_{ijt}] \sim \mathrm{N}\left(\bar{U}_{ijt}, \sigma_U^2\right). \quad (1)$$

Among various formats of information contained in a typical product review system, the arithmetic mean of all numeric ratings submitted is thought to be one of the most straightforward, and perhaps the most influential, piece of information. As a result, we assume the mean of the expected utility to follow:

$$\bar{U}_{ijt} = \bar{R}_{jt} + \lambda_{i0} + \lambda_{i1} \cdot p_{jt}, \quad (2)$$

where $\bar{R}_{jt}$ denotes the arithmetic mean of all numeric ratings of j posted submitted prior to time t, λ_{i0} allows the expected product quality to vary across individuals, β_{i1} captures heterogeneity in price effect on pre-consumption utility formulation, and p_{jt} is the price for the product j listed at time t.

Post-consumption utility. One of the most challenging parts of our modeling framework is to quantify consumers' post-consumption utility. While we can assume submitted ratings are unbiased signals of consumer satisfaction and infer each product experience from the observed ratings, such information is not available if a particular experience does not lead to a review entry (missing data issue). One approach is to assume that the quality of all products follow a normal distribution with a mean of zero (Moe and Schweidel 2012). However, this approach might result in a serious issue; the quality of products with few ratings could be either underestimated or overestimated as the estimation of those products excessively replies on limited data points. To avoid this potential modeling bias, we use long-term mean ratings, $\tilde{R}_j$, as a proxy for baseline product quality.[1] We assume an individual i derives a post-consumption utility U_{ijt} after experiencing the product j purchased at time t in a similar fashion as she derives product expectations:[2]

$$U_{ijt} = \tilde{R}_j + \lambda_{i0} + \lambda_{i1} \cdot p_{jt} \,. \tag{3}$$

It is worth noting that in Equation (3) we generalize the utility construct developed in (Moe and Schweidel 2012) by introducing a covariate p and an individual-specific slope. This specification provides an empirical modeling advantage; it adds an additional degree of heterogeneity into utility formulation and therefore better explains the dispersion of perceived utility across individuals at the product level.

---

[1] In the supplementary data we collected, a product on average has 67 ratings and the minimal number of ratings for a single product is 11. The latest purchase occasion observed in our primary data set occurred in November 2011. A period of over seventeen months, from December 2011 to June 2013, is a reasonable time span for online product ratings to reach the steady state based on the rating submission pattern found in the data.

[2] This simplification is reasonable because of the dual roles a consumer may play while interacting with the online rating system. The same assumption is also used in both analytical (e.g., Sun 2012) and empirical (e.g., Zhao et al. 2013) research in online opinions formation and dispersion.

Utility Disconfirmation. Having developed the pre-consumption and post-consumption utility, we now are able to model the utility disconfirmation. Since in this paper we are interested in the absolute value of the utility disconfirmation, we define

$$\left|\Delta U_{ijt}\right| \equiv \left|U_{ijt} - \bar{U}_{ijt}\right|. \tag{4}$$

Plugging Equation (2) and (3) into (4), the expression for $\left|\Delta U_{ijt}\right|$ is simplified to:

$$\left|\Delta U_{ijt}\right| \equiv \left|\tilde{R}_j - \bar{R}_{jt}\right|. \tag{5}$$

The subscription i is dropped from the right-handed side of the equation becasue $\bar{R}_{\bullet}$ and $\tilde{R}_{\bullet}$ are product- and time-specific only.

5.2 Modeling Updating of the Credibility of the Rating System

Similar to Zhao (2013), we model the credibility of the rating system as the precision with which the post-consumption utility can be reflected by the pre-consumption utility with the same product. We assume that an individual i's prior belief of the system trustworthiness at time t follows a gamma distribution:

$$B_{it} \sim \Gamma\left(a_{it}, b_{it}\right), \tag{6}$$

where a_{it} is the shape parameter and b_{it} is the inverse scale parameter. A consumer will update her belief upon the utility disconfirmation is realized, and as a result, the trustworthiness of the rating system will evolve over time as consumers are involved in more activities of product experience. We assume that a consumer updates her belief in a Bayesian manner (DeGroot 2005) and therefore her posterior belief at time $t+1$ is given by:

$$B_{i,t+1} \sim \Gamma\left(a_{i,t+1}, b_{i,t+1}\right), \tag{7}$$

$$\text{where } a_{i,t+1} = a_{i,t} + \frac{I_{ij,t+1}}{2}, \tag{8}$$

$$b_{i,t+1} = b_{i,t} + \frac{I_{j,t+1}(\Delta U_{ij,t+1})^2}{2(1+1/n_{j,t+1})}, \tag{9}$$

$I_{ij,t+1}$ is a dummy variable indicating whether there is at least one rating submitted for j prior to time t+1, and $n_{j,t+1}$ is the total number of ratings for j available at time t+1.

It is important at this time to point out how the consumer belief evolves over time. Consider an individual i has purchased a product j at time t+1 and experiences j afterwards. If there is no product rating available at time t+1 (i.e. $I_{ij,t+1}$=0), no updating would happen and the individual i's belief remains the same. If product rating signal is available (i.e. $I_{ij,t+1}$=1), however, the extent of updating of the scale parameter b_i is proportional to the utility disconfirmation i perceives as well as the volume of ratings she is exposed to. Such updating rule is realistic because when $n_{j,t+1}$ is small consumers would anticipate the numeric review signal to be noisy with a higher probability. To model how consumers' rating behavior is affected by their idiosyncratic belief, we construct a latent quantity $Cred_{it}$ by calculating the mean of consumers' belief distribution and incorporate it in the raters' decision-making process. Since the belief distribution is modeled as a gamma distribution, the mean of the distribution, denoted by $Cred_{it}$, is given by:

$$Cred_{it} = \frac{a_{it} - 1}{b_{it}}. \tag{10}$$

Since each shopper's complete transaction history is observed in our data set, we are able to construct the up-and-down path of $Cred_{it}$ over time. Finally, we need to estimate each shopper's initial belief. The shape parameter a_{it} governs the richness of an individual's i's belief at time t and the value of a_{i0} indicates the richness of experience i had at the first time he made a purchase on the platform. We fix a_{i0} at a small number across all individuals and assume $b_{i0} \sim N(\overline{b}_0, \sigma_b^2)$.

With this specification, inverse scale parameters are identified at the individual level through the purchase and rating activities observed in the data set.

5.3 Modeling Consumer Decisions on Whether to Rate and What to Rate

Propensity Model. In contrast to the assumption that a review incidence occurs completely at random, we posit that raters' decision on whether to rate is governed by a latent rating propensity. With this stronger economic theory, we are able to model raters' behavior in a more rigorous manner by leveraging those missing rating points. Specifically, an individual *i*'s propensity to rate a product *j* acquired at time *t* is modeled as:

$$Prop_{ijt} = \alpha_i + \beta_1 \cdot |\Delta U_{ijt}| + \beta_2 \cdot Cred_{it} + \beta_3 \cdot U_{ijt} + \beta_4 \cdot U_{ijt}^2 + \beta_p Z_{p,ijt} + \varepsilon_{p,ijt}, \tag{11}$$

where α_i allows for heterogeneity in baseline propensity at individual level, $|\Delta U_{ijt}|$ is the absolute value of utility disconfirmation given in Equation (5), $Cred_{it}$ is the belief construct specified in Equation (10), Z_p is a vector of some control variables such as product price, and finally, $\varepsilon_{p,ijt}$ is an idiosyncratic error with a mean of 0. Moe and Schweidel (2012) find that online raters' rating propensity is affected by the overall ratings environment. To control this external effect, we also include in Z_p some aggregate statistics of ratings environment at product level, such as the number of ratings submitted (volume) and the dispersion of raters' opinion (variance).

The parameters of our main interests are β_1 and β_2. A positive and significant β_1 will suggest that online raters have higher propensity to express their opinions if the discrepancy between their pre-consumption and post-consumption utility are more salient. We can also infer that consumers tend to be silent if the perceived product utility meets their expectation. The estimated β_2 will indicate high credibility of a rating system would encourage or discourage raters to contribute. Moreover, it has been recognized that online opinions is subject to a polar effect – consumers would more likely to express opinions if those opinions are extreme

(Anderson 1998; Dellarocas and Narayan 2006). To model this polar property, we add a quartic term along with a liner term of post-consumption utility into propensity formation. As a result, we can confirm the existence of polar effect if β_3 is negative whereas β_4 is positive.

Evaluation Model. Consider now that a consumer decides to submit a product rating based on the post-consumption utility derived from product experience. We assume that an individual *i* formulates a latent rating evaluation $Eval_{ijt}$, which is the net of post-consumption utility and some other nonutility-related factors:

$$Eval_{ijt} = U_{ijt} + \delta_1 h_{ijt} + \varepsilon_{e,ijt} \text{ ,} \tag{12}$$

, where $\varepsilon_{e,ijt}$ is a zero-mean random shock. While consumers are advised to provide product-related feedback only, we also observe that some consumers also express their dissatisfaction about the merchant's inefficiency in handling process in the textual review body. To account for the possibility that raters may reflect the slow handling time when giving ratings, we include log-transformed handling time window (measured by days) in Equation (12). Since the latent rating evaluation is continuous whereas the score submitted can only be discrete (1~5 in our context), we assume the relationship between these two quantities to follow:

$$z_{ijt} = \begin{cases} 5, & if \;\; \kappa_4 < Eval_{ijt} < \kappa_5, \\ 4, & if \;\; \kappa_3 < Eval_{ijt} \le \kappa_4, \\ 3, & if \;\; \kappa_2 < Eval_{ijt} \le \kappa_3, \\ 2, & if \;\; \kappa_1 < Eval_{ijt} \le \kappa_2, \\ 1, & if \;\; \kappa_0 < Eval_{ijt} \le \kappa_1, \end{cases} \tag{13}$$

where z_{ijk} denotes the score submitted and $\kappa_{i\cdot}$ specifies the valuation-ratings translating cutoffs. For identification purpose we set $\kappa_0 = -\infty$, $\kappa_5 = \infty$ and $\kappa_1 = 0$ (Koop et al. 2007), resulting in three cutoff points κ_2, κ_3 *and* κ_4 to be estimated.

5.4 Modeling Interdependence between Two Rating Decisions

So far we have developed two separate models to describe raters' decisions on whether to rate and what to rate. However, the covariance matrix of the equation system has not yet been clearly specified. Since the observed outcomes for propensity model is binary, we assume $\varepsilon_p \sim \mathrm{N}(0,1)$ and the propensity model simply has a binary probit specification:

$$\Pr(y_{ijk} = 1) = \Phi\left(\alpha_i + \beta_1 \cdot |\Delta U_{ijt}| + \beta_2 \cdot Cred_{it} + \beta_3 \cdot U_{ijt} + \beta_4 \cdot U_{ijt}^2 + \beta_p Z_{p,ijt}\right), \tag{14}$$

where $y_{\cdot}$ is a dummy indicates whether or not a purchase occasion is followed by a review entry. Based on the independence assumption among two decision stages[3], Moe and Schweidel (2012) assume $\varepsilon_e \sim \mathrm{N}(0,1)$ and treat the evaluation model as an independent ordered probit model. However, since the post-consumption utility, U_{ijt}, enters two models simultaneously, the estimation results could be biased if the correlation between two equations are not explicated specified. To empirically test whether raters' decisions are inter-dependent in our data set, we assume that two sets of error terms are governed by a bivariate distribution:

$$\begin{pmatrix} \varepsilon_p \\ \varepsilon_e \end{pmatrix} \sim BVN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right), \tag{15}$$

where the variance of ε_e is fixed at 1 for identification purpose (for evaluation model) and ρ is the correlation coefficient to be estimated. Given this covariance structure and the translating cutpoints defined in Equation (13), the probability of observing a joint event of $y_{ijt} = 1$ and $z_{ijt} = s$ is given by:[4]

[3] 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.

[4] Ying (2006) estimate the correlation coefficient using the inverse Mill ratio as a prediction. The parameter estimations could be inconsistent and the estimated standard errors need to be corrected using certain procedures. We avoid the possible estimation inconsistency by building the likelihood function for data based a conditional bivariate density function.

$$\Pr(y_{ijt}=1, z_{ijt}=s) = \begin{cases} \Phi_2(\infty, Prop_{ijt}, \rho) - \Phi_2(\kappa_4 - Eval_{ijt}, Prop_{ijt}, \rho) & s=5, \\ \Phi_2(\kappa_4 - Eval_{ijt}, Prop_{ijt}, \rho) - \Phi_2(\kappa_3 - Eval_{ijt}, Prop_{ijt}, \rho) & s=4, \\ \Phi_2(\kappa_3 - Eval_{ijt}, Prop_{ijt}, \rho) - \Phi_2(\kappa_2 - Eval_{ijt}, Prop_{ijt}, \rho) & s=3, \\ \Phi_2(\kappa_2 - Eval_{ijt}, Prop_{ijt}, \rho) - \Phi_2(\kappa_1 - Eval_{ijt}, Prop_{ijt}, \rho) & s=2, \\ \Phi_2(\kappa_1 - Eval_{ijt}, Prop_{ijt}, \rho) - \Phi_2(-\infty, Prop_{ijt}, \rho) & s=1, \end{cases} \quad (16)$$

where Φ_2 denotes the standard bivariate normal cumulative distribution function (CDF) and κ_{i1} is fixed at 0 for identification purpose. It is worth noting that, unlike the regular ordered probit specification, the unconditional probabilities of observing five possible scores do not sum to 1 (i.e. $\sum_{s=1}^{5} \Pr(y_{ijt}=1, z_{ijt}=s) \neq 1$). This is because we directly derive the probability of the joint events rather than using conditional probability, and as a result, our proposed model does not rely on the independence assumption between two sets of outcomes (y's and z's). The probability of observing a missing rating point ($y_{ijt} = 0$) simply follows a standard normal density:

$$\Pr(y_{ijk}=0) = 1 - \Phi\left(\alpha_i + \beta_1 \cdot |\Delta U_{ijt}| + \beta_2 \cdot Cred_{it} + \beta_3 \cdot U_{ijt} + \beta_4 \cdot U_{ijt}^2 + \beta_p Z_{p,ijt}\right). \quad (17)$$

Based on Equations (16) and (17), the joint likelihood for individual i who makes m purchase occasions and posts n product ratings is then given by

$$L_i\left(y_{i\cdot}, z_{i\cdot}\right) = \underbrace{\prod_{t \in (y_{ijt}=0)} \Pr\left(y_{ijt}=0\right)}_{(m\text{-}n) \text{ terms}} \cdot \underbrace{\prod_{t \in (y_{ijt}=1)} \Pr\left(y_{ijt}=1, z_{ijk}=s\right)}_{(n) \text{ terms}} \quad (18)$$

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

5.6 Model estimation and model comparison

The methodological novelty of this paper is that we generalize the second stage of the Heckman model to reflect ordinal outcomes (e.g. rating scores) and directly estimate two correlated models without using inverse Mill ratio as a prediction for correlation. We estimate our proposed (full) model using the hierarchical Bayes approach. Based on the deviance information criteria (DIC) of the full model and its three nested versions (see Table 1), we report and interpret the estimation results from the one giving us the best fit (full model).

Table 1. Model Estimation Results

Model	Has individual dynamic feature?	Is $\|\Delta U\|$ included?	DIC (w.r.t. Full Model)
Full	Consumer learning	Yes	---
Model 1	log(t)	Yes	29.3
Model 2	No	Yes	280.5
Model 3	No	No	289.7

6. Results

Table 2 reports the posterior means and posterior standard deviations of parameters of the full model. Since we use Bayesian approach to estimate the model, the significance of parameter estimates is evaluated per 95% highest posterior density (HDP) interval.

Table 2. Model Estimation Results

Model	Parameter	Description	Mean	S.E.	Signif.
Propensity	α_i	Individual-specific propensity intercept	-1.092	0.073	*
	β_1	Utility disconfirmation	0.172	0.045	*
	β_2	Effect of credibility	-0.277	0.021	*
	β_3	Linear term of post-consumption utility	-0.263	0.045	*
	β_4	Quadratic term of post-consumption utility	0.071	0.013	*
	β_5	Product Price	0.278	0.016	*
	β_6	Volume of rating environment	0.010	0.011	
	β_7	Variance of rating environment	0.005	0.030	
	β_8	Utility disconfirmation * Volume	-0.066	0.039	*
Utility	λ_{i0}	Individual-specific baseline utility	2.218	0.074	*
	λ_{i1}	Individual-specific price sensitivity	0.001	0.030	
Evaluation	δ_1	Order handling days	-0.037	0.048	
	κ_{i1}	Cutoff point for s=1 and 2 (fixed at 0)	------	------	
	κ_{i2}	Cutoff point for s=2 and 3	0.498	0.046	*
	κ_{i3}	Cutoff point for s=3 and 4	1.467	0.052	*
	κ_{i4}	Cutoff point for s=4 and 5	2.507	0.053	*

Others	$\log(b_i)$	Individual-specific initial belief status	1.799	0.167	*
	ρ	Correlation coefficient	0.134	0.065	*

* 95% HDP interval does not contain 0

The estimated coefficient β_1 is positive and significant, suggesting that the online shoppers propensity to rate the products is driven by the magnitude of disconfirmation between pre- and post-consumption utility. That is, the larger the extent to which the perceived quality deviates from the *ex-ante* expectation, the more likely an online rater would express her opinions regarding her product experience. From a perspective of opinion sharing, a consumer may think that she has the "responsibility" to let peer consumers know the "true" product quality (from her own standpoint) is actually higher (or lower) than it is rated on the review system. Such attention to voice for the product is even intensified as the degree of disconfirmation increases. On the other hand, a rater's tendency to "correct" the product ratings vanishes as the utility discrepancy decreases. In this case, they don't bother to express opinions if the as-is product ratings precisely signify her perceived quality. This particular finding can be somewhat linked to Anderson and Sullivan (1993) who argue that satisfaction is best specified as the net of perceived quality and quality disconfirmation. The estimated coefficient for disconfirmation-rating volume interaction term, β_8, reveals an interesting story as well. The disconfirmation effect on rating propensity is moderated by the total number of ratings associated with the product. This intrinsic behavioral difference can be linked to the crowding-out effect, a phenomenon where individuals' intrinsic motivation obtained from doing an activity is undermined by some external interventions. While a consumer in general has a desire to correct the ratings, she is discouraged to voice for the product because her contribution is perceived to be relatively marginal if the mean rating represents the opinions from a huge consumer base. A similar pattern is found in a scenario of election where voters void their votes if a particular candidate is declared a clear winner in opinion polls.

Our estimation results also suggest that consumers' rating propensity demonstrates a declining trend ($\beta_2 < 0$) as consumers updating their belief in credibility of the system. This trend implies a consumer is more opinionate when the system credibility is perceived to be low; she would be relatively silent otherwise. One possible explanation is that the more purchases and product evaluation occasions a consumer has made, the less uncertain the quality of purchased products will be, and therefore the less degree of disconfirmation a consumer may have. As for the effect of post-consumption utility on rating propensity, the negative sign of β_3 and the positive sign of β_4 together indicates that consumers are more prone to share opinions when they have either highly satisfied or highly dissatisfied with the product. This finding validates the existence of the polar effect observed by Anderson (1998) and Dellarocas and Narayan (2006). Moreover, the positivity of β_5 suggests that consumers are more interested in voicing for products with higher prices. It is worth noting that the price effect in the propensity model (measured by β_5) and that in the utility model (measured by λ_{i1}) have different economic meanings. The estimate β_5 captures the direct effect of price on rating propensity whereas λ_{i1} governs individual-specific price sensitivity in the utility formation, which affects latent propensity in an indirect fashion. A zero-mean estimate of λ_{i1} also implies that there is no evidence that price can serve as a proxy for product quality. Moreover, the merchant's order handing time is expected to have a negative impact on rating scores ($\delta_1 < 0$). Finally, the significant estimate for ρ suggests that online raters' decisions on whether to rate and what to rate are inter-dependent in our dataset. As a result, two decisions should be jointly estimated to avoid estimation bias.

7. Conclusion

In this research we attempt to identify what factors would affect rating incidence in an online setting. Leveraging a panel data composed of consumers' transactional and review activities, we conclude that disconfirmation between expectation and utility realization is one of the main factors driving review readers to post, perhaps due to the intrinsic responsibility to report "unbiased" ratings. Consumers' attention to voice for products also exhibits a crowding-out effect, meaning that the responsibility of providing feedback is dampened if there has been a big crowd expression their opinions.

This study has few limitations and can be improved in several directions. For example, we think we have not yet extracted full information from our valuable data. In this paper we only investigate consumers' rating decisions and perhaps our framework can be extended to model purchasing decisions as well. In addition, we exclude rating points which are no associated with purchase records in the final data set. It will be also interesting to examine what drives non-buyers to leave feedback.

Reference

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

BIGResearch. 2009. "Word of Mouth Influences Most Electronics, Apparel Purchases, According to Rama Survey." from http://www.nrf.com/modules.php?name=News&op=viewlive&sp_id=845

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

Chintagunta, P.K., Gopinath, S., and Venkataraman, S. 2010. "The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation across Local Markets," *Marketing Science* (29:5), September 1, 2010, pp. 944-957.

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

DeGroot, M.H. 2005. *Optimal Statistical Decisions*. Wiley-Interscience.

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., Zhang, X., and Awad, N.F. 2007. "Exploring the Value of Online Product Reviews in Forecasting Sales: The Case of Motion Pictures," *Journal of Interactive Marketing* (21:4), pp. 23-45.

Duan, W., Gu, B., and Whinston, A.B. 2008. "Do Online Reviews Matter? - an Empirical Investigation of Panel Data," *Decis. Support Syst.* (45:4), pp. 1007-1016.

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].

Kuksov, D., and Xie, Y. 2010. "Pricing, Frills, and Customer Ratings," *Marketing Science* (29:5), September 1, 2010, pp. 925-943.

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.

Li, X., and Hitt, L.M. 2010. "Price Effects in Online Product Reviews: An Analytical Model and Empirical Analysis," *MIS Q.* (34:4), pp. 809-832.

Liu, Y. 2006. "Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue," *Journal of Marketing* (70:3), 2006/07/01, pp. 74-89.

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.

Moe, W.W., and Trusov, M. 2011. "The Value of Social Dynamics in Online Product Ratings Forums," *Journal of Marketing Research* (48:3), 2011/06/01, pp. 444-456.

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

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

Sun, M. 2012. "How Does the Variance of Product Ratings Matter?," *Management Science* (58:4), April 1, 2012, pp. 696-707.

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.

Zhao, Y., Yang, S., Narayan, V., and Zhao, Y. 2013. "Modeling Consumer Learning from Online Product Reviews," *Marketing Science* (32:1), January 1, 2013, pp. 153-169.