REVISITING SELF-SELECTION BIAS IN E-WORD-OF-MOUTH:
AN INTEGRATED MODEL AND BAYESIAN ESTIMATION OF MULTIVARIATE REVIEW BEHAVIORS

Xiao Ma
Wisconsin School of Business
University of Wisconsin–Madison
975 University Ave, Madison, WI
53705
ma29@wisc.edu

Sung S. Kim
Wisconsin School of Business
University of Wisconsin–Madison
975 University Ave, Madison, WI
53705
skim@bus.wisc.edu

Abstract

This paper studies the consumer self-selection bias in the e-word-of-mouth (eWOM) systems, e.g. consumer review websites. Under Bayesian framework, this study extends our understanding of this bias and discovers two new sources through developing a system of structural models of consumer review behaviors tested by a large data set. Our model and results provide evidences that the timing and content of a review introduce significant amount of bias into ratings in a simultaneous fashion. Specifically, we find that after controlling for various exogenous effects the two sources of bias persist: a subsequent rating is positively associated with the time interval between two consecutive reviews by the same consumer, and is negatively associated with the length of a review. Clearly, our findings confirm that modern eWOM systems have notable flaws despite of their mechanical advantages. We further discuss the possible mechanisms as well as the economic impact underlying these findings.

Keywords: E-word-of-mouth, Online consumer reviews, Self-selection bias, Simultaneous equation modeling, Bayesian estimation
1. Introduction

In information systems (IS) and economics literature, e-word-of-mouth (eWOM) refers to the phenomenon that consumers evaluate the quality of products or service providers (henceforth “business”) by providing subjective ratings and reviews through digital channels, primarily in the purpose of informing potential consumers in the future. One popular and successful type of eWOM practices is the online feedback mechanisms (Dellarocas 2003) or reputation systems (Resnick et al. 2000), e.g. in eBay. Since a decade ago, researchers have argued that eWOM, facilitated by the bidirectional communication capabilities of the internet, possesses some mechanical advantages over traditional trust-building mechanisms, especially in large-scale online environment (Dellarocas 2003). An important promise of the eWOM is that potential consumers have access to the previously missing quality assessment of the businesses, thereby mitigating the information asymmetry regarding quality and making the market more effective (Pindyck and Rubinfeld 2008). Based on this promise, online feedback mechanisms have proliferated during the past decade (see a comprehensive overview in Dellarocas 2003).

Along with the rapid growth of the online feedback mechanisms came an expansion of relevant researches especially in recent years (Bodapati 2008; Chevalier and Mayzlin 2006; Chung and Tseng 2010; Clemons et al. 2006; Dellarocas et al. 2010; Duan et al. 2008; He and Chu 2010; Hu et al. 2011; Koh et al. 2010; Li and Hitt 2010). Only a few among these studies have focused on whether the consumer feedback (i.e., ratings and reviews) faithfully reflect the true quality, a question of significant challenge recognized by Dellarocas (2003). Specifically, they have tapped into the consumer self-selection bias in consumer online ratings in a traditional sense (Dellarocas et al. 2010; Hu et al. 2006; Hu et al. 2009; Koh et al. 2010; Li and Hitt 2010; Ying et al. 2006).

Hu et al. (2009) articulate that this bias is primarily of two types: acquisition bias and under-reporting bias; Whereas acquisition bias refers to the scenario that the consumer propensity to buy a product is positively associated with their expected utilities of that product, under-reporting bias depicts the situation where consumers avoid reviewing those products that are neither good nor bad enough. The key notion is that reviews are not generated at random. So far, researchers have only identified this nonrandomness through the consumer’s decision whether or not to generate a review. We identify another two possible sources of bias—both related to the multivariate review decisions—which become prevalent, although previous researches have assumed their nonexistence.

One is the time interval between two consecutive reviews by the same consumer. It falls in the broad topic of the consumer self-selection bias because a consumer decides when to contribute a review; yet such a decision is likely nonrandom. Note that a consumer is unlikely to repeatedly review the same business during a short period. From the perspectives of economics, marketing, and consumer behaviors, this time interval closely relates to the extent of the variety seeking behavior, referring to the mechanisms by which consumers exhibit varied behaviors and tend to switch among similar alternatives (McAlister and Pessemier 1982). Furthermore, previous researches have revealed that variety seeking is an important human characteristic to the success of a company’s investment in customer relationship management, and is attributable to some complex psychological concepts including the consumer satisfaction with a purchase experience (Adjei and Clark 2010; Homburg and Giering 2001; Homburg et al. 2007). Therefore, one of our goals is to examine a positive relationship, if any, between this time interval and a subsequent rating.

The other source—the length of a review—is also important and relevant here because the decision to write a longer review may serve as a catalyst for deliberations and over-criticism. A consumer’s level of deliberative effort when writing a review could directly link to how lengthy the review is. It is also supported that a higher level of deliberations results from the central route of cognitive processes (Petty and Cacioppo 1986, 1996). According to the elaboration likelihood model (ELM), the central route of cognitive processes implies an extensive search for and evaluation of the key product-related features in support of an attitudinal position. Because there generally lacks a solid linkage between a consumer and a business, consumers under the governance of the central route tend to be critical. Therefore, another goal of this study is to examine whether writing a longer review is associated with a lower rating.

To the best of our knowledge, these two sources of the self-selection bias have not been discussed enough. Therefore, this study expects to contribute to the relevant literature in e-commerce in general, and

---

Track 17: Online Communities and Digital Collaborations

2 Thirty Second International Conference on Information Systems, Shanghai 2011
consumer feedback mechanisms in particular, through searching for theoretical frameworks suited to the focal context, as well as developing and empirically testing a system of structural models by a large data set.

2. Literature Review and Research Hypotheses

Most studies of the online feedback mechanisms have been developed within the context of consumer review websites. At the early stage, they primarily focused on the relationship between the aggregated review metrics, e.g. the average rating, and such key economic measures as price and sales of a business (Dellarocas 2003). Later on, new research streams developed, among which one examined the differential impact of these websites under different system designs (Chevalier and Mayzlin 2006). Another expanding stream aims to detect alternative predictors of the economic success of a business other than the average rating. For example, Clemons et al. (2006) find that the heterogeneity in ratings and the level of positivity of those positive reviews well predict the adoptions of a new product. Moreover, Duan et al. (2008) find that review volume instead of the average rating significantly predicts product sales. Duan et al. (2008) also argue that previous researchers have failed to incorporate heterogeneity into their models of consumer review behaviors.

More recently, researchers have delved into the design problems of consumer review websites and made a modeling attempt both in an analytical approach (e.g., Li and Hitt 2010) and through empirical examination. The self-selection issue is by far the most studied. To the best of our knowledge, such an issue is first recognized by Clemons et al. (2006), empirically documented by Hu et al. (2006), and formally modeled by Ying et al. (2006). Moreover, two important focuses of discussions pertaining to such an issue have emerged: its different causes and manifestations, and its impact on ratings (Dellarocas et al. 2010; Hu et al. 2006; Hu et al. 2009; Koh et al. 2010; Li and Hitt 2008; Ying et al. 2006).

On the one hand, several studies have tried to quantitatively understand the various causes and manifestations of this issue. First at a conference (2006) and then in a journal article (2009), Hu et al. articulate that such an issue is primarily of two types: acquisition bias and under-reporting bias; Whereas acquisition bias refers to the scenario that the consumer propensity to buy a product is positively associated with their expected utilities of that product, under-reporting bias depicts the situation where consumers avoid reviewing those products that are neither good nor bad enough. Within the context of the movie industry, Dellarocas et al. (2010) discover that this issue is attributable to the movie’s box office performance. Within the same industry but from a different perspective, Koh et al. (2010) find that people from countries in which social norms emphasize the freedom of expression more than adherence to the socialistic doctrine tend to give very negative ratings and hence are more likely to exhibit self-selection review behaviors.

On the other hand, a few studies have dealt with such a question: to what extent do ratings reflect the true quality? Ying et al. (2006) is the first to contribute to the discussions on this question through rigorous econometric analyses. They show that although it may be difficult to uncover the psychological process by which self-selection decisions are made, it is relatively straightforward to estimate this bias injected into ratings as a result of selective decisions. Li and Hitt (2008) also contribute to the answers to this question. By deriving a later-arrived reviewer’s rating as a function of two key statistics of the earlier ratings, Li and Hitt (2008) find that different combinations of these two statistics are strongly related to quite different sequential patterns of the average ratings across time. Overall, Li and Hitt (2008) shed light on a different source of bias and reaffirm that the self-selection issue in consumer review websites deserves closer examination.

2.1. Research Hypotheses

Since our study inherits several important constructs from Yelp, which is a modern consumer review website from which we collect our data set, it is worthwhile to introduce its basic design features. Yelp is a platform on which consumers review local businesses. Since its foundation in 2004, Yelp has grown significantly. According its official blog, in January 2011 it attracted over 45 million unique visitors. Inside Yelp community, a player has her own homepage, whether she is a consumer/reviewer or business owner/administrator. Yelp offers powerful search utilities to help locate a specific local business. On the
homepage of a business, important economic and business-related information is displayed, including average price per person and the average rating. More importantly, all available reviews for a business are chronologically listed. Similarly, all available reviews contributed by a consumer for different businesses are listed on the homepage of that consumer. When a consumer initiates a review process, she must give a rating that evaluates the service quality of a business using a scale of 1 to 5, as well as write a review to justify the rating. Except for being under public surveillance for creating inappropriate content, a consumer is virtually unlimited when writing a review because it may contain up to 5,000 English words.

In online communities such as Yelp, it is commonly the case that relatively few, peculiarly active users contribute most of the contents ever created. Within the focal context, as long as the evaluations of a business by such a peculiar group of people are not peculiar themselves, the reviews are reliable and unbiased. However, previous researches have already evidenced that these two types of peculiarity tend to occur simultaneously by considerably more than random chance (Dellarocas et al. 2010; Li and Hitt 2008; Ying et al. 2006).

As discussed above, traditional types of the self-selection bias are believed to result from consumers’ nonrandom decisions to contribute reviews. This may be the only valid explanation for nonrandom reviews when IS practices in the consumer services sector were in an exploratory and developing stage. During that time, computers were still “luxury” possessions; inequality primarily existed in whether or not a consumer had access to the internet. However, these have changed dramatically in recent years so that various formats of the online feedback mechanisms are widely accessible. Therefore, traditional types of the self-selection bias are not the only ones in existence. We identify two other likely sources of this bias, neither of which has been fully recognized by previous researches: (1) the time interval between two consecutive reviews by the same consumer; (2) the length of a subsequent review. Besides, we elaborate on the possible links between these two constructs and a subsequent rating.

2.1.1. Time interval between two consecutive reviews

Normally consumers review the exact same product for only once on the same platform. This observation remains true even for non-durable goods because even though consumers are repeatedly engaged in the purchasing experience with a same product, both the product quality and consumers’ evaluation criteria remain relatively stable for a unique consumer-product pair. Such a characteristic in consumer review behaviors is confirmed by our data from Yelp. Therefore, for the same consumer, the time interval between two consecutive reviews for businesses in similar categories indicates the degree of varied behaviors—choices that are different but in similar categories. However, to make such an inference we must assume that, consumers’ decisions to contribute reviews in our data set bear little or only limited amount of the traditional self-selection bias.1

In the economics, marketing, and consumer behavior literatures, a widely discussed type of behavior is closely relevant to the time interval discussed here: variety seeking behavior. Various motives and causes are believed to be involved in variety seeking behavior, which refers to the mechanisms by which consumers exhibit varied behaviors, normally from the perspective of the sequence of choices among similar alternatives (McAlister and Pessemier 1982). An important aspect of variety seeking behavior is the switching of choices, not the uniqueness of the chosen alternatives. In other words, consumers exhibit variety seeking behavior not because of superiority but variation of the choices. Numerous researches have investigated the implication of variety seeking behavior to firms’ investment decisions of customer relationship management (CRM) (Adjei and Clark 2010; Homburg and Giering 2001; Homburg et al. 2007). For example, Homburg and Giering (2001) and Adjei and Clark (2010) find that the relationship between satisfaction and behavioral loyalty is negatively influenced by the tendency to seek for varied behaviors.

However, should variety seeking behavior influence sales, its effect on the consumer satisfaction with the firms has to be examined in the first place, a topic that none of these researches have fully considered.

---

1 Our data set also confirms this assumption to a large part. According to Hu et al. (2009), if the traditional self-selection issue is severe, the marginal distribution of the reviews is bimodal and nonnormal. Such a bimodal distribution of consumer ratings is most documented in product categories such as books, DVDs, videos, and moving pictures (Dellarocas et al. 2010; Hu et al. 2006, 2009; Koh et al. 2010). However, this type of nonnormal distribution is not clearly evident in our data set with most of the reviews for local businesses such as restaurants, supermarkets, and bars.
Indeed, it may well be the case that consumers who exhibit a higher level of variety seeking behavior are less satisfied on average.

A consumer’s preference can be thought of as a single peaked function of product attributes (McAlister and Pessemier 1982). As the amount of an attribute cumulates with repeated experiences, its positive marginal effect on the preference utility decreases below zero, until some point in time when she is motivated to switch choices in order to maintain a high level of utility. This point in time is referred to as the “point of satiation” (Coombs and Avrunin 1977). When a consumer feels urgent in switching to another alternative, her marginal utility is likely to already fall below zero and keeps dropping substantially. However, because consumers are usually confined to a limited set of alternatives due to geographical immobility or inertia in the human behavior, they tend to be disappointed with the extent of variation even if an altered choice is made successfully; thus, they become less satisfied. On the contrary, as time goes by the influence from previous experiences wanes; a consumer who waits long enough before contributing another review will not likely be subject to this disappointment. Therefore, we expect that the shorter this time interval, the lower a subsequent rating is. H1 is proposed as follows,

H1: When a consumer rates and reviews a business online, the time interval between two consecutive reviews is positively associated with a subsequent rating.

2.1.2. Length of a subsequent review

It is a widely adopted tactic that consumer review websites encourage consumers to write reviews besides giving a rating. The rich information contained in these reviews has recently attracted an exploding stream of studies that falls into the discussions of data mining (Chung and Tseng 2010; Decker and Trusov 2010; Hu et al. 2011; Mackiewicz 2010; Tsang and Prendergast 2009). Although the reviews available online can be quite informative for corporate marketers to infer specific consumer preferences, writing such a review provides no tangible benefits to self and sometimes can be energy-consuming. It has been argued that generating contents online in exchange of no tangible benefits remarkably resembles the gift-giving behavior in interpersonal communications: Both of them are primarily motivated by the intent to build or maintain one’s social status, whether in real life or in online communities (Bergquist and Ljungberg 2001; Lampel and Bhalla 2007).

When writing a review to justify a rating, consumers are unlimited regarding how long a review could be. This allows for enormous amount of consumer heterogeneity in how elaborated a consumer decides to be. We posit that an appropriate framework to help understand the implication of this level of elaboration is the elaboration likelihood model (ELM) (Petty and Cacioppo 1986, 1996; Petty et al. 1983), which argues that humans adopt two different routes when processing information: they tend to use a central—or “logical”—route when they are more involved, whereas they adopt a peripheral—or “heuristic”—route when they are less involved. ELM has been widely applied to understand broad human subjects including communications, persuasion, and learning.

At earlier times when consumer review websites are less sophisticated and customers are only required to give ratings, they tend to adopt the peripheral route. Later as they are allowed to do more, such as crafting a qualitative review, they may be more involved into the evaluation process and thus tend to switch to a central route more often. The rationale is that, to justify a rating, a consumer needs to more effortfully search for convincing evidences in the memory of the purchasing experience and then write a well-structured easy-to-read review. Motivated by this situational cue, a consumer becomes more involved and the central route is thus more likely to be activated. Because the central route of processing information requires a consumer to base her evaluations solely on concrete evidences, and because there generally lacks a solid linkage between a consumer and a business, she tends to ignore social norms and her general beliefs; instead, she tends to focus on the actual service performance of a business, sometimes even trivial and irritating details. Therefore, H2 is proposed the follows,

H2: When a consumer rates and reviews a business online, the length of a subsequent review is negatively associated with a subsequent rating.
3. Theoretical Analysis and Model Development

We note that a key modeling concern is closely relevant here: the multivariate and simultaneous review behaviors. Researchers nowadays are able to observe various types of consumer behaviors (e.g., product purchases, customer feedback, website visits and bookmarking) as their “footprints” are easily recorded by the modern computer technologies. These “footprints” provide excellent observations to help understand consumer preferences (Decker and Trusov 2010; Ying et al. 2006). Meanwhile, these behaviors tend to interact with each other in a dynamic and simultaneous fashion. Arguably, there are few good reasons against a modeling attempt that could integrate as many aspects of consumer behaviors as possible, apart from the difficulty in doing so. Moreover, the benefits of such an attempt have already been demonstrated within the same (Ying et al. 2006) and other e-commerce contexts (e.g. online auctions, e.g. Park and Bradlow 2005). Therefore, we construct a system of three structural modules to simultaneously examine the following consumer review behaviors: (1) the time interval between the previous and subsequent reviews (henceforth “When” module), (2) how lengthy a subsequent review is (“Text” module), and (3) what a subsequent rating is (“Rating” module).

3.1. “When” Module

The “time interval” type of data has been the central focus of the models on the adoption and diffusion of various types of products (Bass 2004; Sawhney and Eliashberg 1996; Xue, Hitt, and Chen 2011). The key to these statistical models is that, a time interval indicates both an adoption decision and the speed (e.g., sooner or later) thereof. In other words, earlier adopters presumably have higher utilities regarding the adopted product than later adopters. By the same token, a time interval between two consecutive reviews by the same consumer indicates their unique preferences. Time interval data generally follow an exponential distribution with a long tail to the right. This is a situation that violates the normal distribution assumption of the dependent variable in classical linear regression (CLR) models (Greene 2011). Thus, we transform this time interval variable by taking its natural logarithm. The transformed time intervals approximately follow a normal distribution.

\[
\ln(W_{i,j,t}) = \sum_{a}^A \beta_a^w W_{i,j,t}^a + \sum_{b}^B \beta_b^w W_{i,j,t}^b + \sum_{c}^C \beta_c^w W_{i,j,t}^c + \sum_{d}^D \beta_d^w W_{i,j,t}^d + \eta_{W_{i,j,t}} \sim N(0,\sigma^2_w) \quad (1)
\]

Equation (1) illustrates how we model time intervals represented by \(W_{i,j,t}\), where \(i\) index a consumer, \(j\) index a business, \(t\) index a review occasion, and \(a\) through \(d\) denote different exogenous variables of the corresponding types. These exogenous variables include time-varying consumer behaviors \(s_{(W_{i,j,t})}\), consumer-specific demographics \((X_{(W_{i,j})})\), business-specific characteristics \((Y_{(W_{j})})\), and time-specific environmental variables \((Z_{(W_{i,j})})\). \(\eta_{W_{i,j,t}}\) is a normally distributed stochastic term that captures the unobserved factors (e.g., dynamic endogenous shocks) and measurement errors. After controlling for various exogenous variables, a consumer’s idiosyncratic preference pertaining to the time interval, which is hard to measure objectively, may persistent and remain in \(\eta_{W_{i,j,t}}\).

3.2. “Text” Module

We present our specification for “Text,” represented by \(T_{i,j,t}\), module in Equations (2):

\[
\ln(T_{i,j,t}) = \sum_{a}^A \beta_a^t X_{(T_{i,j,t})}^a + \sum_{b}^B \beta_b^t Y_{(T_{i,j})}^b + \sum_{c}^C \beta_c^t Z_{(T_{i,j})}^c + \sum_{d}^D \beta_d^t Z_{(T_{i,j})}^d + \eta_{T_{i,j,t}} \sim N(0,\sigma^2_t) \quad (2)
\]

Where the scripts \(i, j, t\), and \(a\) through \(d\) inherit their previous definitions. After considering all of these exogenous variables, a consumer’s idiosyncratic preference pertaining to the length of a subsequent review remains in \(\eta_{T_{i,j,t}}\).

3.3. “Rating” Module

Perhaps the most prevalent mechanism for consumer ratings uses an integer value that ranges from 1 to 5 to assess the overall quality of a product or service. Econometric modelers have argued that ordinal and
censored data such as ratings require different model specifications from those that are appropriate for normally distributed data (Ying et al. 2006). Therefore, we follow the flexible approach detailed in Ying et al. (2006) when we parameterize the ratings, as described in Equation system (3):

\[
\begin{align*}
Pr(U \leq \kappa_k) & \quad \text{If } k = 1 \\
Pr(Y = k) = & \begin{cases} 
Pr(\kappa_{k-1} < U \leq \kappa_k) & \text{If } k \in [2, K - 1] \text{ and } k \in N \\
Pr(\kappa_{k-1} \leq U) & \text{If } k = K
\end{cases}
\end{align*}
\]  

Where \( k \) is the possible ratings that range from 1 to \( K \); \( \kappa_k \) through \( \kappa_k \) are cutoff points similar to those in Ying et al. (2006); \( U \) is a random variable that denotes consumers’ latent utility, or their anticipated satisfaction with a business \((U_{ijt})\). It is further specified as in Equation (4):

\[
U_{ijt} = \sum_{a}^{A} \beta_{a}^{R} R_{ijt} + \sum_{b}^{B} \beta_{b}^{Y} Y_{ijt} + \sum_{c}^{C} \beta_{c}^{x} x_{ijt} + \sum_{d}^{D} \beta_{d}^{s} s_{ijt} + \mu_{W_{ijt}} \delta_{w,R} + \mu_{T_{ijt}} \delta_{r,R} + \epsilon_{(R)ijt} \quad \text{Wherein } \epsilon_{(R)ijt} \sim \mathcal{N}(0,1)
\] 

For identification purposes, we follow Ying et al. (2006) and estimate \( K - 2 \) cutoffs after mandatorily letting \( \kappa_1 \) be negative infinity and \( \kappa_K \) be positive infinity. Note that \( \epsilon_{(R)ijt} \) is a normally distributed stochastic term whose standard deviation is set to 1 for identification purposes (Wooldridge 2002). Finally, to appropriately assess the bias of a rating associated with the time interval and length of a subsequent review, respectively, \( \mu_{W_{ijt}} \) and \( \mu_{T_{ijt}} \) are included in Equation (4), and \( \delta_{w,R} \) and \( \delta_{r,R} \) are used to represent their coefficients.

4. Data

Data collection process includes the following steps. We first select a large snowball pool of 26,360 consumers from registered members of Yelp in January 2011, from which a random subsample is generated, whose history of review behaviors are then tracked. We make sure that each of these consumers had contributed at least two reviews. This resulted in a large data set of 118,538 review occasions of a total of 2,420 consumers.

Because of the page limit, we briefly introduce and summarize the descriptive statistics of the important research variables. Most importantly, When, Text, and Rating, respectively, is the dependent variable of each of the three modules. Mean and standard deviation of Rating, respectively, is 3.74 and 1.10. The variables in \( s_{ijt} \) include: (1) LagUtility, which is inferred from the previous rating and represents the latent utility that determines the previous rating, and (2) CumuExpr, which is a natural log-transformed cumulative number of reviews contributed so far in the history of a consumer. \( x_{ijt} \) are reviewer demographic variables such as dummy variable Gender, log-transformed number of Friends,\(^2\) and GeoMobil, which stands for Geographical Mobility and is the proportion of reviews for out-of-state businesses. Variables included in \( y_{ijt} \) are the most relevant business-specific characteristics: four dummy variables \((\text{Category1} \text{ to Category4})\) that indicate the consumer services type of a business;\(^3\) Price that indicates approximate cost per person; AvgRate which is the arithmetic mean of the ratings available for a business; and a dummy variable FreeWiFi that indicates whether a business provides free wireless internet connection. Finally, we include several control variables, as included in \( z_{ijt} \), such as a dummy variable (Facebook) that indicates whether a review was contributed before or after Yelp incorporated the friends’ network of Facebook; and a total of five dummy variables \((\text{Year2005 to Year2010})\) that control for year-specific unobservable shocks. Therefore, to better test for the existence

\(^2\) Consumers may form dyadic relationship with another consumer within the network of Yelp.

\(^3\) These dummy variables indicate the following types of consumer services, respectively: (1) food or leisure, (2) travel or lodging, (3) shopping, and (4) legal, educational, healthcare, or professional; therefore, the fifth category with all of these four dummies being zero include all other types, such as non-governmental and non-profit organizations.
and estimate the influence of two newly identified types of the self-selection bias, we control for many important exogenous variables among which previous researches only include a partial list (Ansari et al. 2000; Ying et al. 2006).

5. Estimation and Preliminary Results

Bayesian estimation techniques have been widely used in medical and natural science fields in the 80’s and 90’s. More recently they are applied to management and social science fields including marketing (Rossi et al. 2006). Under hierarchically structured models, Bayesian methods offer much flexibility and are well suited for problems in marketing and IS, such as IS adoption behaviors. In essence, these types of research problems require estimations at both within-individual and across-individual levels. Based on the prior knowledge of unknown parameters, hierarchical Bayesian methods first estimate within-individual coefficients which is independent across individuals; then, coefficients at higher levels are estimated based on individual-specific parameter estimates (see more detailed discussions on the advantages of Bayesian methods in Rossi et al. 2006).

Therefore, we adopt the Bayesian framework and utilize the Markov chain Monte Carlo (MCMC) simulation method in estimating our structural models. This approach provides at least two advantages: (1) At each iteration of the simulation process, we update \( \hat{\mu}_{W_{i,j,t}} \) and \( \hat{\mu}_{T_{i,j,t}} \) in Equation (4), respectively, using their consistent estimators from Equations (1) and (2), the \( \hat{\mu}_{W_{i,j,t}} \) and \( \hat{\mu}_{T_{i,j,t}} \). This helps ensure that the estimations of the “Rating” module are consistent and efficient (Stephen and Toubia 2010). (2) To effectively control for individual heterogeneity, random effects are incorporated into each of the coefficient parameters for the time-variant research variables (those in \( s, x, y, \) and \( z \)). Simulation-based estimation methods, such as the one implemented here, offers the feasibility in estimating the model under current specifications that optimization-based analytic tools lack, such as the maximum likelihood estimation (MLE) based on Laplace approximation.\(^4\)

Other than the iterative updating of the two bias-related terms, we follow the standard procedure of a hierarchical Bayes specification, including an improved mixture of Gibbs and asymptotically efficient Random-Walk Metropolis algorithms (Rossi et al 2006). As a preliminary analysis, we calibrate our model on the data of 241 consumers. This subset of data bears similar characteristics to the complete data set. The simulation chain ran for 50,000 iterations as the burn-in period and another 50,000 for estimating the posterior distributions (Ntzoufras 2009; Rossi et al. 2006). Due to a lack of memory of our computing devices and also for reducing auto correlations of the MCMC chain (Ansari et al. 2000; Ying et al. 2006), we use the process to preserve every 5th of the 50,000 posterior draws. All the chains seem to quickly converge to the equilibrium distributions. The random-walk Metropolis algorithm performs well, because the chains are stable and the overall average acceptance rate stabilizes at about 0.28 which is very close to its optimal level (Ntzoufras 2009; Rossi et al. 2006).

To test our hypotheses, we shall focus on the estimations for \( \delta_{W,R} \) and \( \delta_{T,R} \). \( H_1 \) is supported: if the time interval increases by one unit, the latent satisfaction that determines the rating increases by .025 (\( p < .001 \)). \( H_2 \) is also supported: if the length of a review increases by one unit, the latent satisfaction decreases by .04 (\( p < .001 \)).

6. Discussion, Implication, and Future Plan

To conclude, this study continues the discussions on the biased design of the online feedback mechanisms. We aim to identify and test for the existence of two additional sources of the self-selection bias that previous researches have not fully elaborated on. Specifically, we posit that given the modern design features of consumer review websites and the specific context of Yelp, a popular US-based consumer review website devoted to providing evaluations of local businesses, decisions of the time interval between two consecutive reviews by the same consumer and decisions of the length of a review

\(^4\) We verified that our model with current specifications had a hard time converging to an optimal state in either the Mixed model in SAS or the nlme-based package “ordinal” for ordinal response models in R.
Ma & Kim / Revisiting Consumer Self-Selection Bias in eWOM Systems

simultaneously introduce significant amount of the self-selection bias into ratings. Since more research is even needed to confirm what we find here, it is beyond the mission of our study to fully uncover the processes by which they arise; instead, we apply the plausible theoretical framework in trying to understand the processes.

First of all, our results confirm the postulation that as the time interval between two consecutive reviews decreases, the subsequent rating also decreases. Although we reason that such a downward bias as a result of shortened time intervals is attributable to a consumer’s intensified variety seeking behavior, other alternative explanations could exist. For example, consumers may review more businesses during a short time when their friends temporarily visit them. To appear “smart” in a group, their friends tend to criticize the businesses and eventually they are influenced by their friends’ negative comments. However, the results indeed imply that this association is not by random chances. Modern consumer review websites are advised to correct for the bias resulting from time intervals that are too short. For example, for a business that has not received a significant amount of reviews (e.g., 100 reviews), even a little bias in consumer ratings is critical and biased ratings tend to influence its future sales. The websites may examine if any of the reviews for this business are by any consumer who was temporarily very active so that the review was contributed immediately after (within the same day or only 1 or 2 days apart) reviewing another business. If so, her rating can be assigned a weight less than 1 before being used to produce the average rating for that business, as a way to undermine the possible bias introduced from an extremely small time interval.

Second, our results support a negative association between the length of a subsequent review and the subsequent rating. As such, shorter reviews tend to introduce an upward bias while longer reviews introduce a downward bias into ratings. Although the intent of consumer review websites to encourage consumers to contribute more elaborated reviews is well justified, the payoff may not be satisfactory. When those socially active online content creators are given the opportunity to elaborate their evaluations, they may tend to pay more attention to trivial features and incidents and eventually become overcritical. Extremely long reviews are indeed not necessary because given many reviews, the average rating will slowly but ultimately converge to the true quality, as long as traditional types of the consumer self-selection issue is not severe. Given the sheer scale of the popular consumer feedback mechanisms nowadays, practitioners need to focus on correcting for this bias from review lengths for those new or niche businesses that have received only a small number of reviews (e.g., fewer than 100 reviews).

Although tentative, our model and results send out a warning signal that the expanding scale of the online feedback mechanisms makes ratings less capable of reflecting the true quality. Therefore, such an expanding scale produces significantly more “smog” that blurs consumers’ vision and makes product improvement more difficult.

To carry on this study and overcome the limitations pertaining to the current theoretical framework, we are working on better explaining the mechanisms of the newly discovered types of the bias. We will examine more closely the endogeneity issue, a critical assumption that our system of simultaneous models is built on. We will also enhance our model to better account for consumer heterogeneity, which has been emphasized as a necessary component in dealing with consumer review behaviors (Ying et al. 2006).

Even though we have assumed that traditional types of the self-selection bias is not severe in our context and have demonstrated the face validity of this assumption, it is better to rigorously test and control for it, if any. We will calibrate our revised model on multiple random subsamples to assess how robust it is. Moreover, we are also working on demonstrating the flexibility of our modeling approach and discovering possible ways of comparing the magnitude of the self-selection bias with those previously uncovered.

Acknowledgements

We would like to thank Yun Deng, an exchange Ph.D. student from the School of Management, Huazhong University of Science and Technology, for his great help during the data collection process.
References


