Simulation Modelling Practice and Theory 77 (2017) 108–123

Contents lists available at ScienceDirect

Simulation Modelling Practice and Theory

journal homepage: www.elsevier.com/locate/simpat

The effectiveness of online reviews in the presence of self-selection bias

Bharat Bhole*, Bríd Hanna*

Department of Economics, Rochester Institute of Technology, Rochester, NY 14623, USA

ARTICLE INFO

Article history:
Received 3 January 2017
Revised 25 May 2017
Accepted 26 May 2017

Keywords:
Online reviews
Word-of-mouth
Self-selection bias
Under-reporting bias
Mean star-rating

ABSTRACT

Online reviews suffer from self-selection biases. One of these is under-reporting bias: those who have an extreme experience, either positive or negative, are more likely to review a product than those who have a moderate experience, and consequently are overrepresented in the sample of reviews. In this paper we study whether under-reporting bias decreases the effectiveness of the mean star-rating as a criterion for choosing between competing products of different but unobservable qualities. We formulate a model of consumer choice and the decision to review, and simulate this model in a variety of scenarios. We find that if under-reporting bias decreases the effectiveness of the mean star-rating, this is usually only for products that are of either extremely poor or extremely good quality. Even in these cases, the effectiveness is reduced only when the extent of bias is extreme and the variance in quality is low. Otherwise, the presence of under-reporting bias generally does not decrease, and in fact, often enhances the effectiveness of the mean star-rating as a measure of relative quality.

1. Introduction

There is now substantial evidence that online reviews play an important role in consumer choice. According to the Local Consumer Review Survey by Brightlocal [3], about 84% of the respondents trusted online reviews as much as a personal recommendation; and 91% of respondents read reviews (50% regularly and 41% occasionally) to determine if a local business is good or bad. According to the Local Search and Online Reviews Survey by ReviewTrackers [16], 62.7% of the respondents considered online reviews "important" or "very important" when choosing a local business; and while 63.29% of respondents trusted businesses with 4.0 stars or higher, only 2.5% trusted businesses with lower than 2.0 stars. The importance of online reviews found in these surveys is confirmed by many academic studies based on actual reviews and sales data (as opposed to survey data). For example, Chevalier and Mayzlin [4], Chintagunta et al. [5] and Ye et al. [19] all find in different contexts that review valence (online star-rating) has a positive and significant effect on sales.[1] A meta-analysis by Floyd et al. [9] based on a study of 443 valence elasticities finds the average valence elasticity of sales to be 0.69.

Given the significant influence of online reviews on consumer choice, it is important to know whether they are effective in communicating the relative quality of the products. Specifically, does the use of mean star-ratings cause consumers to

* Corresponding authors.
E-mail addresses: bharat.bhole@rit.edu (B. Bhole), bxggse@rit.edu (B. Hanna).

[1] The Chevalier and Mayzlin [4] study is based on reviews of books on Amazon.com and BarnesandNoble.com; the Ye et al. [19] study is based on online reviews of hotel stays in China; and the Chintagunta et al. [5] study uses box-office ticket sales data and data on movie reviews from the Yahoo! Movies website.

http://dx.doi.org/10.1016/j.simpat.2017.05.005
1569-190X/

make correct choices?[2] This question becomes especially important given that online reviews are beset with certain problems: one, they are subject to different types of self-selection biases (these biases are discussed below), and two, they can contain fake reviews. The objective of this paper is to examine whether one of the self-selection biases, namely the under-reporting bias, has an adverse influence on the effectiveness of mean star-ratings.[3]

The under-reporting bias refers to the phenomenon that consumers who have a more extreme experience (either positive or negative) are more likely to review a product or engage in word-of-mouth communication about the product than those with a moderate experience [1,2,11–13].[4] Using different datasets from the Amazon.com website, both Hu et al. [11] and Lafky [13] find that moderate ratings (especially, 2 or 3 star ratings) are less common than extreme ratings (1 star and 5 star ratings), even for products that have a moderate average rating. One possible explanation that has been suggested by Hu et al. [12] and Lafky [13] for the presence of under-reporting bias is that rating a product requires time and effort and hence consumers are unlikely to rate a product unless they have a sufficiently strong incentive to do so. Further, consumers are more likely to have a stronger incentive to rate a product when they feel more strongly about the product (either positively or negatively). This causes people with a more extreme experience to be overrepresented in the sample of reviews.[5]

There is also some experimental evidence in favor of under-reporting bias. In a controlled experiment carried out by Hu et al. [12], the authors find that when all consumers of a music CD had to review the product (and, hence there was no scope for under-reporting), the distribution of the reviews was unimodal. On the other hand, the review distribution for the very same product on Amazon.com was J-shaped. Lafky [13] also presents experimental evidence of polarization in ratings when rating is costly.[6]

Assuming that the available ratings are characterized by under-reporting bias, the question we are interested in is: does under-reporting bias decrease the effectiveness of valence, specifically the mean star-rating, as an indicator of the product's relative quality. For example, Hu et al. [11] argue that in the presence of under-reporting bias the mean star-rating does not reveal a product's true quality and can be misleading.[7] We study this question by simulating a model in which consumers choose between two alternatives, named H and L, on the basis of their prior beliefs and the mean star-rating.[8] Product H is of higher quality in the sense that the distribution of payoff (consumer's satisfaction) from H first order stochastically dominates that from L. Consumers are, however, unaware of this fact. We use the final market share of product H (denoted by α_H) as an indicator of the effectiveness of the mean star-rating. It is socially optimal that every consumer chooses product H. Hence, the higher the market share of H the greater is the effectiveness of reviews. To investigate the influence of bias on the effectiveness of the mean star-rating we compare the market shares of H with and without under-reporting bias (denoted by α_H^B and α_H^U respectively), while controlling for the number of reviews.

While it is true that in the presence of under-reporting bias the mean star-rating is not an unbiased estimate of the true mean quality of the product, our analysis suggests that this generally does not imply a reduction in the effectiveness of the mean star-rating as a measure of a product's relative quality. We find that if under-reporting bias decreases the effectiveness of the mean star-rating, this is usually only for products that are of either extremely poor or extremely good quality. Even in these cases the effectiveness is reduced only when the extent of bias is extreme and the variance in quality is low. Otherwise, the presence of under-reporting bias generally does not decrease, and in fact, often enhances the effectiveness of the mean star-rating as a measure of relative quality.

Gao et al. [10] also study, amongst other things, whether online ratings are a good indicator of quality. They find strong and consistent evidence of a positive relationship between quality and online ratings.[9] Further, similar to our findings, Gao et al. [10] suggest that online ratings are more informative for physicians in moderate quality segments, but are less effective in differentiating quality among "high-end" physicians. The likely reason for this according to Gao et al. [10] is the hyperbole effect of online raters, which is tendency to exaggerate experiences of extreme quality, either good or bad. Our findings show that extreme under-reporting bias can also be responsible for reducing the effectiveness of online ratings when products are

[2] Based on several consumer studies designed to assess how consumers use ratings and other observable cues to form quality inferences, De Langhe et al. [6] find that "[Consumers] place enormous weight on the average user rating as an indicator of objective quality compared to other cues. Averages based on small samples and distributions with high variance are treated the same as averages based on large samples and distributions with low variance."

[3] Features of online reviews other than the mean star-rating can also affect consumer choice. For example, the volume of reviews, the variance of star-ratings, characteristics of reviewers, and the verbal content of reviews can all have an influence on consumer decisions (see De Maeyer [7], Dellarocas et al. [8], Sun [17], Ye et al. [19]). However, we have not incorporated these influences in this paper. Our objective is to focus on how the effectiveness of valence is influenced by the presence of under-reporting bias, and our results can be interpreted as those obtained when holding these other features of reviews constant. Given the importance of valence, and the strong evidence in favor of the under-reporting bias, we believe this is a worthwhile first step.

[4] A different type of self-selection bias identified by Li and Hitt [14] is one where the early consumers of a product are likely to be favorably disposed to the product, and hence are more likely to give it positive ratings.

[5] Another possible explanation for a preponderance of extreme ratings is the presence of fake reviews (see Luca and Zervas [15]). In this paper we focus on under-reporting bias and assume that consumers review the products truthfully.

[6] Fig. 6 in Lafky [13] shows the review distribution for both free and costly treatments. Lafky notes "... in the face of a small cost of rating, people are more willing to rate when they have either a very positive or very negative experience relative to a more moderate one."

[7] Hu et al. [11] say that "if the average score reveals a product's true quality, a consumer would be better off by purchasing a product with an average score = 3 instead of one with an average score = 2, other things being equal. However, when the average score does not reveal a product's true quality in the traditional sense, this is not necessarily true."

[8] We also run some simulations with three products to check the robustness of our results. Our main findings hold even in this case. See Appendix B.5.

[9] Gao et al. [10] use data from a survey of a representative sample of patients as the objective measure of physician quality.