

A literature review, submit to Rotterdam School of Management:

Building an Online Reputation with Free Content

Siming Yan, syan52@wisc.edu

Feb 11, 2020

Summary:

This article focuses on figuring whether a **free content provision** can help e-book content providers to build up an online reputation and further become successful.

Background:

The previous literature shows that: A good online reputation makes a seller successful. However, new sellers face the challenge of a cold start, since the hardness of building a reputation over the products that already gained traction in the market. Whereas, free content provision can be a useful tool for sellers to gain reviews or online reputations, either good or bad.

Therefore, to make a new seller easier to obtain his or her success, the author needs to answer the following questions. How to examine the true effect of the free content provision on reputation and success in the e-book market? What are the relationships between free content provision and reputation, unilateral or bilateral? Are these impacts positive, or negative while comparing to those e-books at a certain positive price?

To solve the questions above, the author divides the impacts of free content provision into two aspects, across-consumer and within-consumer. On the one hand, for across-consumer effects, previous literature proves that there exist self-selection effects: lower preference consumers induced by the free price, tend to rate the e-book worse since they comparably don't like it initially. On the other hand, for within-consumer effects, some say the same consumer should rate better when viewing it as a gift; others say the reviews could be worse, as higher price often means higher quality. It seems that we cannot easily draw any conclusions without performing some empirical modeling.

Data Description:

To build up empirical models, the author collects two complementary datasets. One is a panel dataset that contains 4 single months of e-book data across 4 years (2015-2018) on *Smashwords*, a leading self-published e-book distributor website. The data contains characteristics of the books, numbers of reviews and average rating, and author traits. This dataset is used for testing the relationship between free content provision and reputation. The other dataset is taken from the last wave of the data and contains detailed ratings and review contents. The later dataset is used for identifying the features of the impacts of providing free content provision.

Methods:

The methods when identifying the relationships between free content provision and reputation include linear regressions with e-book, platform, fixed effects, and time-varying controls. Also, the linear probability model and conditional fixed effects logit model are

introduced to check if rating can increase the probability of a content to be set as free. The methods applied on the review content level additionally include text-based sentiment model and lasso-method to choose word phrases and trigrams to avoid overfitting. The author also uses linear regression predicted rating as the dependent variable to summarize the net effect of selection. The author also creates some variables, such as *match characteristics* to improve the accuracy of models.

Results:

Based on the datasets and methods above, the author obtains the following results:

1. Free content receives 0.05 to 0.06 lower ratings (scales 0 to 5) compares to when it is offered at a positive price. This negative effect can be explained by the selection effect: zero price attracts people with a lower preference to read and rate the book.
2. The review contents, the length of the characters, and generated sentiment score are less or lower when the content is offered for free. This result is generated by Natural Language Processing with lasso-regression.
3. The rating distribution on free content is 5 to 6 percent more dispersed when compared to paid content. This is marked by the absolute deviation of a book rating to the overall average rating of the same book.
4. On the reviewer's characteristics side, female reviewers, reviewers with previous same author reviews, or higher mean ratings, have on average higher ratings on the new content. However, this effect drops by half in size when the e-book is free.
5. The free content provision can improve or decrease the rating of an e-book varies from sub-genres to sub-genres. The standard deviances are different as well.

To summarize the main results in a take-home note: free content provision makes a seller gain more early-stage reviews and reputation, as a tradeoff of decreased and unstable rating and worse sentiment review contents. However, this effect can vary from one sub-genre to another.

These results are somehow compatible with what Lin et al. (2019) have found from the free sample provision impacts on an e-commerce platform but extend to the e-book market. In both works, free content provision, and free sample provision are proven to effectively increase the number of reviews. However, due to the methods of delivering free sample provision are different, the attitudes of reviews are different. In Lin's work, reciprocating a free product is introduced to attract more positive reviews and positive ratings. In this work, free digital content provision has approximately no produce and distribute cost for the content provider, therefore the reviews are comparably more negative and scattered.

Questionable Assumptions:

The article has shown a concisely deductive procedure for developing the theories. However, the basis of the theories is established on several assumptions:

Firstly, these three months of data are randomly selected and can be a plausible sample of all data in the e-book market. I think using more data and cover a longer period can be more plausible.

Secondly, when investigating the relationship between free content and rating, the author assumes the effect of selection is the main reason account for the rating decrease. This is arguable, although the author used linear model prediction to identify the selection effect as true. The results of controlling for observed characteristics of reviewers, the remained effect of offering content for free is statistically insignificant. A given explanation is across-consumer selection dominates within-consumer effects. However, there could be other explanations, for example, the free contents are less worthy reading, but when paying for it, consumers are uneasy to admit their failure in choosing. Or, the book could be modified after moving from free to some price, the price is for the cost of the modification. A possible solution to these potential explanations is, tracking the modified history and the word count has been modified.

Another questionable thing is, what is the length of the book that has been read when the consumer submits the review and rating? The length has been read might have significant impacts on rating. However, this is untraceable due to the limitation of the website. I have noticed that apps such as WeChat reading do have this function, which could be useful for further studies.

Conclusion:

I think the paper is overall highly rigorous and logical. The inference part is strictly developed, from a seller's perspective, to test a possible way to success step by step. The models are precisely defined, and the natural language processing part stunningly reflects the quality of the reviews. This is an article with thorough thoughts and reached to any aspect of the free content provision impacts on reviews, rating, and content providers' success.

Reference:

Lin, Z., Zhang, Y., & Tan, Y. (2019). An empirical study of free product sampling and rating bias. *Information Systems Research*, 30(1), 260–275. <https://doi.org/10.1287/isre.2018.0801>