

Zegners, Dainis

Conference Paper

Building an Online Reputation with Free Content: Evidence from the E-book Market

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2017: Alternative Geld- und Finanzarchitekturen - Session: Empirical Industrial Organization I, No. B12-V2

Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

Suggested Citation: Zegners, Dainis (2017) : Building an Online Reputation with Free Content: Evidence from the E-book Market, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2017: Alternative Geld- und Finanzarchitekturen - Session: Empirical Industrial Organization I, No. B12-V2, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, Kiel, Hamburg

This Version is available at:

<https://hdl.handle.net/10419/168293>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Building an Online Reputation with Free Content: Evidence from the E-book Market*

Dainis Zēgners

LMU Munich, Institute for Strategy, Technology and Organization (ISTO)

March 1, 2017

Current Version:

http://dainiszegners.net/ebooks_reputation

Abstract

An important strategy to build a reputation in markets with asymmetric information is to practice introductory pricing. By selling their product at a lower introductory price, sellers can increase demand, induce more buyers to provide feedback, and thus build a reputation more quickly. I examine a form of introductory pricing that is particularly popular in digital markets, namely offering free content to consumers. I argue that offering free content to build a reputation can be a double-edged strategy. It does not only attract buyers with a high preference, but also buyers with a low preference for the product. Low-preference buyers give worse feedback, resulting in a negative selection effect on a seller's reputation. I estimate the strength of this effect using data from an online self-publishing platform where I observe independent e-book authors either selling their e-books at a price or giving them away as free content. By exploiting the fact that I observe ratings for free and purchased versions of the same e-book, I show that those consumers who obtain an e-book as free content rate it worse than reviewers who buy it at a positive price, consistent with a negative selection effect on reputation.

Keywords: Online Reputation, Ratings, Free Content

JEL codes: L86, M31, D82

*I thank my colleagues at the ISTO, my Ph.D. advisors Tobias Kretschmer and Florian Englmaier, Tim Schweisfurth, Sofronis Clerides, audiences at the *TIME Colloquium* in Munich, the *EEA 2016* conference in Geneva, the *EARIE 2016* conference in Lisbon, the *INFORMS 2016* conference in Nashville, the *Media Economics Workshop 2016* at UZH Zürich, and seminar audiences at University of Surrey and Télécom ParisTech for many helpful suggestions and discussions. Contact details: Institute for Strategy, Technology and Organization (ISTO), Munich School of Management, Ludwig-Maximilians-University Munich, Kaulbachstr. 45/II, D-80539 München, e-mail: d.zegners@lmu.de.

1 Introduction

In markets with asymmetric information, it is crucial for sellers to establish a reputation. Especially in online markets, where often sellers are anonymous, reputation systems such as online ratings are important to reduce information asymmetries between buyers and sellers (Dellarocas, 2003; Cabral, 2012). However, in establishing a reputation, sellers may face a start-up problem: Without a reputation, it can be hard to sell, but only by selling and buyers providing feedback a seller can build a reputation. Strategies to solve this problem are to practice introductory pricing to increase sales (Villas-Boas, 2004), to provide consumers with free samples (Bawa and Shoemaker, 2004; Peitz and Waelbroeck, 2006), or to give away products for free. These strategies contribute to building a reputation by increasing the number of buyers who experience the seller's product and thus the information about the product that is passed on to other potential buyers via word-of-mouth or online ratings. Once the seller has established a good reputation, he can monetize his reputation by enjoying higher sales or earning a price-premium (Shapiro, 1983; Ba and Pavlou, 2002; Resnick and Zeckhauser, 2002).

In this study, I focus on the strategy of giving away products for free to increase buyer feedback. I argue that theoretically this strategy can have either a positive or a negative effect on a seller's reputation: First, giving away products for free may create a negative selection effect on reputation. Free products may not only attract buyers with a high preference, who would also be willing to buy the product at a positive price, but also buyers with a low preference, who are only willing to acquire the product if it is offered to them for free. These low-preference buyers might give worse feedback, decreasing the average feedback that the seller receives, resulting in a worse reputation. Second, receiving a free product may induce buyers to give better feedback, as they might reciprocate the seller's gift of a free product (Kahneman et al., 1986; Ra-

bin, 1993; Falk and Fischbacher, 2006). Alternatively, buyers might provide feedback according to their indirect utility, i.e. utility minus the product's price. As receiving a product for free, instead of paying a price, increases buyers' indirect utility, providing "better value for money," buyers might give better feedback for a free product.¹ Both mechanisms, acting reciprocal or rating according to indirect utility, would result in better average feedback for free products and thus in a better reputation for the seller.

I test for the presence of these effects in the context of readers providing online ratings for e-books on an online self-publishing platform. On this platform, authors regularly give away their e-books as free content. The platform actively encourages this strategy, arguing that it helps authors to establish a readership and increase the number of online ratings their e-books receive. A particular feature of the platform is that it provides the information on its website whether the reader assigning a given rating received an e-book as free content or purchased it at a positive price. Since a subset of e-books has received ratings for both free and purchased versions, I can employ a fixed effects regression approach using only within e-book variation in ratings to estimate the effect of offering an e-book for free on ratings. Thereby, I can test for the presence of the hypothesized effects only by looking at whether those buyers who receive an e-book as free content give it a better or worse rating compared to buyers who purchase the same e-book at a positive price.

My results confirm the presence of a negative selection effect: An e-book that is offered as free content receives on average a rating that is 8 to 10 percent of a standard deviation lower than the rating of a purchased version of the same e-book. Moreover, ratings for free e-books have on average a 5 to 12 percent higher standard deviation of ratings than purchased versions. Furthermore, these effects are stronger for e-books in genres where the standard deviation of ratings is generally higher, consistent with a stronger selection effect for "niche" products, for which buyer valuations are more dispersed. I also find that reviewers who obtain a free copy of an e-book write reviews that are between 6 to 8 percent shorter. However, I also find some evidence for reciprocity

¹Or according to a Russian proverb: "Even vinegar tastes sweet if it's obtained for free."

in the form of the negative selection effect being substantially weaker for female reviewers, which is in line with previous research in Behavioral Economics showing that women generally act more reciprocal than men (Croson and Buchan, 1999).

These results imply that giving away products for free is a risky strategy. Although it increases the probability of feedback and makes it easier for a seller to build a reputation, it also decreases the average feedback while simultaneously increasing its dispersion, resulting in a worse reputation of the seller. An implication of the negative selection effect is that sellers should target their free products at buyers they know have a high preference for their products, for example, based on previous purchases or other characteristics of buyers observable to sellers.

My study contributes to the following strands of literature: In recent years, a growing literature both in economics and marketing has studied online reputation and online reputation systems. Dellarocas (2003) and Cabral (2012) survey this literature. Many studies use data collected from *eBay* to establish the relevance of online ratings for sellers' success in the marketplace. For example, Cabral and Hortacsu (2010) find that a negative *eBay* rating decreases a seller's weekly sales growth rate from a positive 5 percent to a negative 8 percent. Other studies have established that sellers with a good online reputation in the form of a high average online rating earn a price premium (Dewan and Hsu, 2004). The relevance of online ratings has also been shown in other contexts: For example, Claussen et al. (2013) show that a better rating increases the number of active users of Facebook apps, Farajallah et al. (2016) show that car drivers with better ratings can more easily share rides on a peer-to-peer drive-sharing platform, or Chevalier and Mayzlin (2006) show that good ratings increase sales for books on e-commerce platforms such as *Amazon*.

Building on these findings establishing the importance of a seller's online reputation, another literature stream examines how sellers try to influence or manipulate their online ratings: Dellarocas (2006) finds evidence for fraudulent sellers rating their own products or paying third parties to rate their products. Friedman et al. (2007) show that sellers are more likely to change their online identities after receiving bad

feedback. In a similar vein, Wibral (2015) confirms in a lab setting that if sellers can remain anonymous by changing their online identity without buyers observing this, buyers have overall less trust in sellers. Seller anonymity has, therefore, adverse effects on the amount of trade in a market. Nosko and Tadelis (2015) show in a field experiment conducted on *eBay* that a platform can increase buyers' participation in the market by providing additional measures of a seller's quality than only showing his average rating. There is also strong evidence that reciprocal behavior is important within online reputation systems, as in many settings buyers cannot only rate sellers, but sellers can also rate buyers. Examples of such bi-directional reputation systems can be found on platforms such as *eBay*, *AirBnB* or *Uber*. Overall, there is strong empirical support for reciprocity behavior in the context of online ratings (e.g. Bolton et al., 2013; Klein et al., forthcoming).

Within the literature on online reputation, my study is particularly close to Li and Xiao (2014) and Cabral and Li (2015). Both articles propose to incentivize buyers to give feedback by offering them rebates on their purchasing price in case they provide feedback. Li and Xiao show in a lab setting that offering rebates conditional on giving feedback increases buyers' propensity to provide feedback. Yet, Cabral and Li find little support for this hypothesis in a field experiment conducted on *eBay*. However, they find evidence for buyer reciprocity, as buyers give better feedback in the form of a higher rating if they are offered a rebate. My study departs from these articles by proposing and studying an additional selection effect: Offering buyers a rebate conditional on providing feedback is, in principle, equivalent to lowering the product's price, given that the rebate is larger than buyers' cost of providing feedback and that buyers know ex-ante about the rebate. But a lower price should also induce buyers with a lower preference for the product to purchase it. If lower-preference buyers provide worse feedback, offering a rebate or lowering the price should induce a negative selection effect on a seller's reputation. Offering a product for free as in my empirical setting can be thought of as an extreme form of lowering the price.

A possible explanation why both articles do not find a negative selection effect is

that in the case of the lab setting in Li and Xiao (2014), participants were induced to have homogeneous preferences, and in Cabral and Li (2015) the authors intentionally chose to sell a homogeneous product (USB sticks) on *eBay*. Additionally, buyers on *eBay* usually do not rate the purchased product per se but rather a seller's complementary services such as an accurate description and timely shipping. For such services, buyers are likely to have comparatively homogeneous preferences. Therefore, a negative selection effect is likely to be small on *eBay*. In the empirical context of my study, however, I study a setting with products for which buyers should have comparatively heterogeneous preferences in the form of a strong personal taste component, namely e-books. As an e-book can be considered a typical case of an experience good (Nelson, 1970), an author's reputation is also likely to be particularly important in this setting.

My study can also be related to an extant theoretical literature in economics studying reputation. See Cripps (2009); Mailath and Samuelson (2006); MacLeod (2007); Cabral (2005); and Bar-Isaac and Tadelis (2008) for literature reviews. In this literature, reputation is understood as the expectation of buyers regarding future outcomes connected to a seller. Such an expectation can be either based on an inherent type of a seller that buyers learn over time, in the simplest case whether a seller's product is of a "good" or "bad" quality, or whether the seller can credibly commit to behaving in a certain way, in the simplest case whether he will exert "high" or "low" effort. Models examining the second type of reputation based on a seller's commitment typically analyze the long-term equilibria of infinitely repeated games, abstracting from dynamic considerations but focusing instead on "bootstrap" equilibria. They are, therefore, less suitable to analyze dynamic aspects of reputational concerns such as reputation building (Bar-Isaac, 2004, p.16). Models of the first type, modeling reputation as buyers' belief of the seller's inherent type, are better suited for modeling reputation building. In particular, Bergemann and Välimäki (2000, 2006) study the incentives of competing sellers to encourage experimentation and learning of their products' quality by setting a low price. Other similar models are McFadden and Train (1996), and Bose et al. (2006). However, these articles typically assume homogeneous preferences of buyers

regarding the type of the seller. Two recent exceptions are Bar-Isaac and Deb (2014b,a), who, however, do not consider pricing in their models. Therefore, my study highlights a trade-off that to-date is absent from the theoretical literature on reputation.²

My study is structured as follows: In section 2, I develop a short theoretical framework and derive testable propositions. In section 3, I describe the dataset I use to test these propositions. In section 4, I test these propositions with my dataset. In section 5, I conclude.

2 Theoretical Framework

In this section, I present a short theoretical framework to derive propositions that I empirically test in the next sections. The framework addresses two related questions: First, how does giving away products for free affect the nature feedback, i.e. whether buyers give positive or negative feedback? Second, how does giving away products for free affect the dispersion of feedback?

Assume a market with a single product sold by a monopolistic seller for which buyer i has utility

$$u_i = q + \epsilon_i,$$

where q is the product's quality valued equally by all buyers, and ϵ_i is a buyer's idiosyncratic taste for the product, which is distributed according to a continuous random variable with expectation $E(\epsilon_i) = 0$. Before buying, the buyer does not know his utility u_i for the product but he can inspect the product to receive a noisy signal of his utility given by

²There is also a small theoretical literature that looks at similar issues. For example, Villas-Boas (2004) studies a model where consumers are more likely to repeat purchases of experience goods that they have purchased in a first period, which gives firms the incentive to aggressively price in a first period in order to build market share. Another example is Campbell et al. (2013), who study in a model of word-of-mouth firms' incentives to strategically limit the amount of word-of-mouth, as buyers might prefer to signal their status by consuming products that fewer people are informed about (e.g. indie music).

$$S(u_i) = u_i + \theta_i,$$

where θ_i is a noise term that is distributed according to a continuous random variable with expectation $E(\theta_i) = 0$. The buyer then buys the product if

$$S(u_i) > p,$$

where p is the product's price.

After buying the product, a buyer provides feedback in the form of a rating with a constant probability s . The probability that any given buyer i gives feedback is then

$$\rho_i = s \times \Pr(q + \epsilon_i + \theta_i > p),$$

which depends both on a buyer's probability to give a rating s , and on the probability that a buyer buys the product, which he does if the sum of his personal taste component ϵ_i , the product's quality q , and the realization of the noise term θ_i is larger than the product's price p . Assuming that buyer i posts a rating r_i corresponding to his utility u_i ,³ which he learns perfectly after consuming the product, the expected value of an observed rating is

$$E(r_i | S(u_i) > p) = q + E(\epsilon_i | q + \epsilon_i + \theta_i > p),$$

which does not only depend on the product's quality and a buyer's taste but is also conditioned on actually observing a rating.

The seller faces the following trade-off when trying to increase the probability of feedback: Either he can lower his price p , increasing probability ρ_i that buyer i buys the product and provides feedback, but at the same time lowering the conditional expectation $E(r_i | S(u_i) > p)$ of a rating, or he can increase his price p , decreasing probability

³This assumption is common in the literature, c.f. Cabral and Hortacsu (2010); Cabral and Li (2015); and Sun (2012).

ρ_i that buyer i rates the product, but increasing the conditional expectation of a rating. How the seller resolves this trade-off depends on how sensitive future demand is to the availability and nature (good or bad) of feedback.

In the empirical context of my study, I compare ratings of e-books that are offered either at a positive price ($p > 0$) or at a zero price ($p = 0$). As the conditional expectation of an observed rating in equation (2) is increasing in price, i.e.

$$\frac{\partial E(r_i | S(u_i) > p)}{\partial p} > 0,$$

this implies the following proposition:

Proposition 1. *An observed rating of a free product will on average be lower than the rating of a purchased product (ceteris paribus).*

In contrast to this proposition, buyers might also rate a free product better than a product that they purchased at a positive price. There might be two reasons: Either buyers act reciprocal, repaying the seller's gift of a free product with a better rating, or they rate a product according to their indirect utility. Regarding reciprocity, an established literature in Behavioral Economics has shown that people often act reciprocal, meaning they are eager to repay favors, gifts, and invitations etc. (Kahneman et al., 1986; Rabin, 1993; Falk and Fischbacher, 2006). There is also an established literature observing reciprocal behavior within online reputation systems (e.g. Bolton et al., 2013; Cabral and Li, 2015; Klein et al., forthcoming). If a product is given away for free, reciprocal buyers might reciprocate the sellers "gift" of a free product by assigning it a better rating. An observationally equivalent mechanism might be that buyers do not assign a rating according to their utility but according to their indirect utility, i.e. $r_i = u_i - p$. If this is the case, buyers should rate free products more favorably, as a free product represents "better value for money." In my empirical context, both mechanisms are empirically indistinguishable.⁴ Both can be captured within my framework by observed ratings

⁴One way to empirically distinguish between both is to look at whether receiving a product for free makes it more likely that a buyer assigns a rating. If a buyer is not reciprocal but simply assigns a rating

having a conditional expected value of

$$E(r_i|S(u_i) > p) = q + E(\epsilon_i|q + \epsilon_i + \theta_i > p) - p.$$

If both mechanisms are stronger than the selection effect, the following relationship between price and rating should be observed:

$$\frac{\partial E(r_i|S(u_i) > p)}{\partial p} < 0,$$

implying the following alternative to the previous proposition:

Proposition 1A: *An observed rating of a free product will on average be higher than the rating of a purchased product (ceteris paribus).*

An additional testable implication of the selection effect is that a higher price, implying a stronger selection effect, should increase the dispersion or variance of observed ratings, i.e.

$$\begin{aligned} \text{Var}(r_i|S(u_i) > p) &= q + \text{Var}(\epsilon_i|q + \epsilon_i + \theta_i > p) \\ \implies \frac{\partial \text{Var}(r_i|u_i > p)}{\partial p} &< 0. \end{aligned}$$

This implies the following proposition:

Proposition 2: *Observed ratings of free products will on average have a higher variance than the ratings of purchased products (ceteris paribus).*

3 Empirical Setting

To test the three propositions I have derived in the previous section, I use data collected from the online self-publishing platform *Smashwords.com*. While initially a niche-market,

according to his indirect utility, receiving a product for free should increase the rating he assigns but not the likelihood that he assigns a rating. If a buyer, on the other hand, is reciprocal, receiving a product for free should increase both his probability of assigning a rating and how good the ratings is, as a reciprocal buyer should be more eager to reciprocate the seller's gift. Cabral and Li (2015) are not able to find a statistically significant effect of offering a rebate on buyers' probability to provide a rating.

direct self-publishing of e-books by independent authors on platforms such as *Smashwords* has evolved in recent years to a market of substantial size. This growth was mainly triggered by a decrease in production and distribution costs to virtually zero with the advent of digital publishing over the Internet. By 2013, the market share of self-published e-books has grown to an estimated tenth of both the number of books in bestseller lists and overall unit sales (Waldfogel and Reimers, 2015).

Smashwords is one of the largest online distributors of self-published e-books (Bowker, 2014). E-books that are published on *Smashwords* are not only sold on its website but also distributed to the largest online e-book retailers such as *Apple's iBookstore*, *Barnes & Noble*, or *Kobo*. Although most titles on *Smashwords* are written by amateur authors, who serve only small niche markets, some titles published on *Smashwords* have also been very successful, even reaching international bestseller status (Coker, 2012). For more information on self-publishing and *Smashwords*, see Zegners (2016).

The following features make the independent e-book market a good setting to examine the strategy of giving away free products with the goal of establishing a reputation: Whereas authors who release their books via traditional publishers are supported by publishers running advertising campaigns, inducing newspapers to publish reviews, or organizing book tours, self-published authors cannot make use of these complementary services provided by publishers. To inform potential readers of their books' existence, characteristics, and quality; self-published authors have to rely mostly on social media, word-of-mouth, and online ratings. Offering e-books for free is a particularly popular strategy to reach these goals. Often, authors either temporarily offer e-books for free by setting their price to zero, or they offer certain e-books permanently for free, for example, the first volume of a series of books. *Smashwords* actively encourages this strategy with the following arguments:

For many first-time authors, one of your biggest challenges is to get your first readers and your first reviews at retailers. Free builds readership and can help you establish your first reviews . . . Or, if your book has been out a long time and sales have dwindled, try a temporary free promotion to rev up readership, reviews and word of mouth.

Source: <http://blog.Smashwords.com/2015/08/how-to-price-kindle-books-to-free.html>.
(Accessed on February 15, 2016).

Smashwords also regularly runs promotional campaigns where it offers e-books of enrolling authors for free.

3.1 Dataset

I collected the data I use in this study during February 2016 directly from *Smashwords*'s website using a web crawler.⁵ Overall, my dataset contains more than 300,000 e-books offered on *Smashwords*. More than 45,000 e-books have been rated by at least one verified buyer.⁶ To test whether I find the theoretical propositions confirmed with this data, I use a particular feature of the *Smashwords* website: For each rating, the website indicates whether the rating has been given for a free e-book or a purchased e-book. See figure 1 for a screenshot of an e-book's website on *Smashwords*, showing an example of this information attached to each rating.

For my analysis, I use a subset consisting of 45,634 e-books on *Smashwords* that have been rated at least once. Overall, these e-books have been rated 109,421 times. Table 3 shows descriptives of the variables on the e-book level. On average, an e-book has been rated 2.40 times, with an average rating of 4.42 stars (buyers can rate an e-book with 1 to 5 stars). While the average price of an e-book is \$1.94, at the time where I collected the data 17,790 of rated e-books (39 percent of all rated e-books) were offered for free.

For 24,927 e-books, all ratings were given for a free copy. For 18,483 e-books, all ratings were given for a purchased version. For 2,199 e-books, ratings were given for both a purchased and a free version of the same e-book. While initially I include all

⁵In another article (Zegners, 2016), I use a snapshot collected from *Smashwords* in September 2015 to examine how offering part of an e-book as a free sample (not the full e-book as in this study) impacts an e-book's price. I do not include the variable indicating whether a sample is offered for an e-book in this study, as a sample can only be offered for a non-free e-book.

⁶Only verified buyers can rate a non-free e-book on *Smashwords*.

Table 1: Description - Variables and Controls on the E-book Level

Variable	Description
Free E-book (Yes=1)	Dummy indicating whether an e-book's price is equal to zero
Price	E-book's price in U.S. Dollars
N Ratings	Number of ratings an e-book has received on <i>Smashwords</i>
Average Rating	Average rating of an e-book on <i>Smashwords</i>
Length E-book in Words	Total length of e-book in words
Time Since Published (Weeks)	Number of weeks passed since an e-book has been published
Category	Categorical variable indicating category (genre) of e-book
Subcategory	Categorical variable indicating subcategory (subgenre) of e-book
Language	Categorical variable indicating language of e-book
Year Published	Categorical variable indicating year e-book has been published on <i>Smashwords</i>
Gender Author	Probability that author is female as implied by first name
Number E-books	Total number of e-books same author has published on <i>Smashwords</i>
Number Previous E-books	Number of e-books same author has previously published on <i>Smashwords</i>
N Ratings Previous E-books	Number of ratings that are available on <i>Smashwords</i> for previous e-books of same author
Average Rating Previous E-books	Average rating on <i>Smashwords</i> for previous e-books of same author

Notes: This table describes the variables contained in my dataset that are available on the level of each e-book.

Table 2: Description - Variables and Controls on the Rating Level

Variable	Description
Rating (1-5 Stars)	Rating a reviewer assigned to an e-book
Reviewed after Published (Weeks)	Number of weeks between date rating was assigned and date e-book was published on <i>Smashwords</i>
Free E-book Rated (Yes=1)	Dummy indicating whether a rating was given for a free e-book
N Previous Ratings	Number of previous ratings assigned to an e-book before a given rating
Gender Reviewer	Probability that reviewer is female (as implied by first name)
Characters in Review	Number of characters contained in written review

Notes: This table describes the variables contained in my dataset that are available on the level of each individual rating.

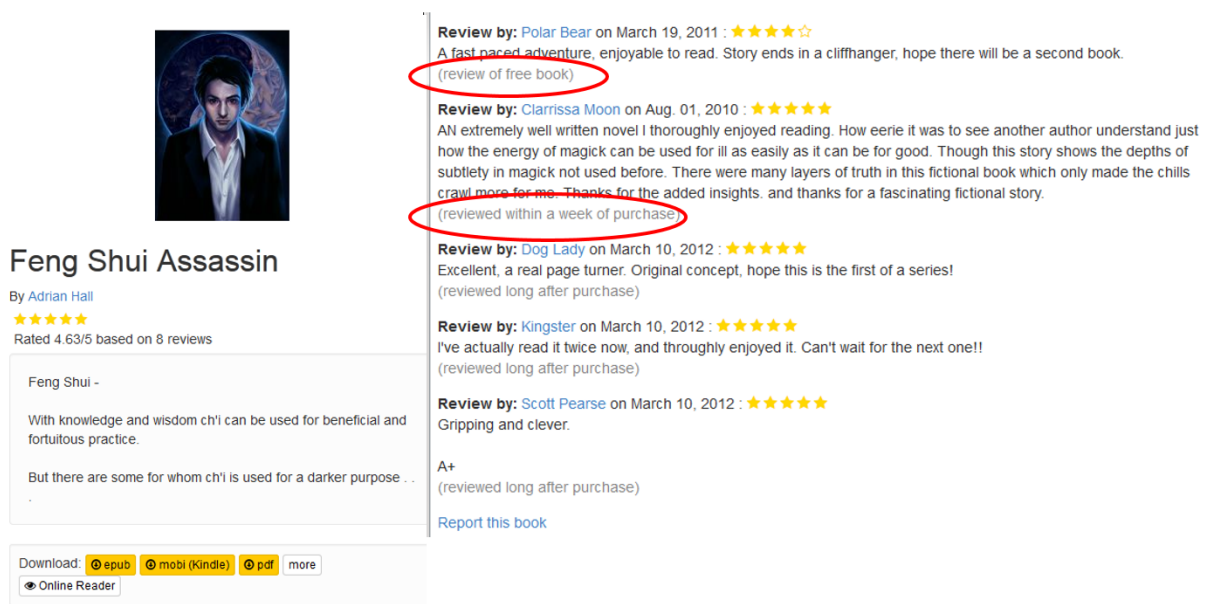


Figure 1: *An E-book's Website on Smashwords*

Notes: This figure shows a screenshot of an e-book's website on *Smashwords* from which I scraped my dataset. The e-book's website shows, additional to information such as the genre and subgenre of the e-book, ratings posted by previous readers of the e-book. Additional to the rating assigned (1-5 stars) and a review, the website shows whether the rating was given for a free version ("review of free book") or for a non-free version ("reviewed within . . . [time]. . . purchase").

Table 3: *Summary Statistics - Variables on E-book Level*

Statistic	N	Mean	St. Dev.	Min	Median	Max
Free E-book (Yes=1)	45,609	0.39	0.49	0	0	1
Price	45,609	1.94	5.79	0	0.99	999
N Ratings	45,609	2.40	4.67	1	1	422
Average Rating	45,609	4.42	0.83	1	5	5
Length E-book in Words	45,609	42,845	48,694	10	23,570	1,234,310
Time Since Published (Weeks)	45,609	173.60	72.56	20.86	175.60	404.30
Gender Author	45,609	0.52	0.44	0	0.49	1
Number E-books	45,609	17.17	165.30	1	6	17,242
Number Previous E-books	45,609	5.99	52.85	0	1	7,284
N Ratings Previous E-books	45,609	3.38	18.67	0	0	1,072
Average Rating Previous E-books	22,763	4.42	0.64	1	4.58	5

Notes: This table shows summary statistics of the variables on the e-book level contained in my dataset. Only e-books with at least one rating are included.

ratings into my analysis, ultimately the subset of e-books where I observe ratings for both free and purchased versions will be central to my identification strategy.

Table 4 shows descriptives of the variables on the rating level contained in my dataset. The average of ratings is 4.48 and on average a written review posted together with a rating contains 488 characters. On average, a review is published 188 weeks after an e-book has been first published on *Smashwords*. About 45 percent of ratings are assigned to free e-books, while the remaining 55 percent of ratings are assigned to non-free e-books. Additional to these variables, I estimate the probability of a reviewer being female as implied by his or her first name, using a database based on U.S. Social Security Administration baby name data implemented in the R-package *gender* (Mullen, 2015).

3.2 Empirical Relevance of Free E-books

Before testing my main propositions, I present some descriptive and cross-sectional e-book level evidence showing the empirical relevance of free e-books. My goal is not to offer a rigorous analysis but to show that the cross-sectional evidence is consistent with the underlying theoretical motivation and mechanisms that I propose.

Table 5 shows a comparison of the means of the e-book level variables contained in my dataset, splitting the sample into free and non-free e-books. All means of these variables are statistically different between both groups. While non-free e-books are rated on average 2.22 times, free e-books are rated on average 2.68 times. This indicates that giving away e-books as free content indeed might help to increase the number of ratings that an e-book receives. On the other hand, while non-free e-books receive on average a rating of 4.51 stars, free e-books receive on average a rating of 4.29 stars. This difference could be either caused by the negative selection effect I propose, or by authors of lower-quality e-books being more likely to offer their e-books for free. Free e-books are also on average shorter and are more often written by male authors. Consistent with offering e-books for free being used as a strategy by new authors to establish a reputation, the authors of free e-books on average have a lower number

of e-books offered on *Smashwords* and have gathered on average fewer ratings. Also, authors of free e-books have worse average ratings for their other e-books compared to authors of non-free e-books.

Table 7 shows a cross-sectional logit regression estimating the probability that a given e-book is offered for free. Additional to including all variable on the e-book level, I also include category, subcategory, and language fixed effects into the regression model. All coefficients are statistically significant with effects in the same direction as suggested by the comparison of means between free and non-free books.

So overall, looking at the cross-sectional evidence, offering e-books for free seems to be indeed a strategy used by entering authors. However, as free e-books are on average almost twice as short as non-free e-books, which might impact their ratings, selection into free and non-free has to be taken seriously in any empirical analysis.

4 Results

In this section, I empirically test the propositions laid out in section 2. First, I test propositions 1 and 1A by examining whether free e-books receive better or worse ratings. Second, I test proposition 2 by examining whether the dispersion of ratings is higher for free e-books. Last, I provide evidence for the presence of a negative selection effect by testing three additional mechanisms: Whether free e-books receive longer or shorter reviews, whether there are differences in the effect size within different sub-genres of e-books, and whether the information on previous e-books a reviewer has rated can provide further evidence in favor of a negative selection effect on ratings.

As already mentioned, I use the fact that *Smashwords* provides the information whether the reviewer assigning a rating received a free version or purchased the e-book at a price. Although in each analysis, I also show results from cross-sectional regressions, my main identification strategy is to use e-books where I observe ratings for free and non-free versions of the same e-book.

Table 4: *Summary Statistics - Variables on Rating Level*

Statistic	N	Mean	St. Dev.	Min	Median	Max
Rating (1-5 Stars)	109,421	4.48	0.87	1	5	5
Free E-book Rated (Yes=1)	109,421	0.45	0.50	0	0	1
Reviewed After Published (Weeks)	109,421	189.20	74.20	20.86	192.70	404.30
N Previous Ratings	109,421	5.25	21.23	0	1	421
Gender Reviewer	109,421	0.61	0.38	0	0.61	1
Characters in Review	109,421	488.10	708.50	1	244	22,301

Notes: This table shows summary statistics of the variables on the rating level contained in my dataset.

Table 5: *Comparison of Groups - Free and Non-free E-books*

	Non-free E-books N=27,819	Free E-books N=17,790	<i>p</i> -value (H_0 : Equal Means)
N Ratings	2.22 (3.71)	2.68 (5.85)	<0.001
Average Rating	4.51 (0.77)	4.29 (0.90)	<0.001
Length E-book in Words	52,586 (49,710)	27,612 (42,831)	<0.001
Time Since Published (Weeks)	182 (70.7)	160 (73.4)	<0.001
Gender Author	0.55 (0.44)	0.47 (0.43)	<0.001
Number E-books	19.3 (210)	13.9 (31.4)	<0.001
Number Previous E-books	6.44 (66.1)	5.28 (18.0)	0.005
N Ratings Previous E-books	3.75 (21.0)	2.79 (14.2)	<0.001
Average Rating Previous E-books	4.47 (0.60)	4.33 (0.70)	<0.001

Notes: This table shows a comparison of means of the variables on the e-book level contained in my dataset, splitting the sample into free (price equal zero) and non-free e-books (price above zero).

Table 7: Probability E-book is Offered for Free

	<i>Dependent variable:</i>
	Free E-book (Yes=1)
	<i>Logit</i>
Log(1 + N Ratings)	0.016*** (0.001)
Average Rating	−0.059*** (0.003)
Log(Length E-book in Words)	0.000*** (0.000)
Time Since Published (Weeks)	−0.001*** (0.000)
Gender Author (Female=1)	−0.068*** (0.005)
Log(Number E-books)	0.009*** (0.003)
Log(1 + Number Previous E-books)	−0.040*** (0.003)
Log(1 + N Ratings Previous E-books)	−0.017*** (0.005)
Ratings Previous E-books Available (Yes=1)	0.043*** (0.015)
Average Rating Previous E-books (Demeaned)	−0.050*** (0.005)
Constant	0.751*** (0.040)
Category (Categorical)	Yes
Subcategory (Categorical)	Yes
Language (Categorical)	Yes
Observations	45,609

*p<0.1; **p<0.05; ***p<0.01
Coefficients give marginal effects estimates
Standard errors in parentheses

Notes: This table shows the results of a logit regression with the dummy indicating whether an e-book is offered for free as the dependent variable. Marginal effects are calculated for each observation and then averaged. The variable *Average Rating Previous E-books* is demeaned and filled up with zeros in the case no rating is available for an e-book or a previous e-book.

Table 8: Mean and Standard Deviations of Ratings for Free and Non-free E-books

<i>Sample: All E-books</i>	μ	σ	N
Rating Non-free E-book	4.564	0.768	60,016
Rating Free E-book	4.382	0.963	49,405
Difference	0.182	-0.194	
p -value (H_0 : Equal Means or Equal Variance)	< 0.001	< 0.001	
 <i>Sample: E-books with Ratings from Free and Non-Free Versions</i>			
	μ	σ	N
Rating Non-free E-book	4.434	0.816	8,902
Rating Free E-book	4.459	0.887	5,710
Difference	-0.025	-0.071	
p -value (H_0 : Equal Means or Equal Variance)	0.08	< 0.001	
 <i>Sample: E-books with Ratings for Free and Non-free Versions</i>			
	μ	σ	N
Rating Demeaned Non-free E-book	0.030	0.653	8,902
Rating Demeaned Free E-book	-0.047	0.726	5,710
Difference	0.076	-0.073	
p -value (H_0 : Equal Means or Equal Variance)	< 0.001	< 0.001	

Notes: This table shows means and standard deviations of ratings for different subsets of e-books. In the top panel, all ratings are used, while in the middle and bottom panel only ratings are used of e-books where both ratings for free and non-free versions are available. Demeaned ratings in the bottom panel are calculated by subtracting from each individual rating the average rating of the e-book that the rating was assigned to.

4.1 Descriptive Evidence

Table 8 provides first descriptive evidence by showing average ratings and the standard deviation of ratings, splitting the sample into ratings provided for free and for non-free e-books. All ratings are pooled, ignoring that ratings are assigned to different e-books. The upper panel of table 8 shows average ratings for the whole sample. The average rating of e-books for non-free versions is higher than for free versions ($p\text{-value} < 0.001$), providing first evidence in support of a negative selection effect (Proposition 1) and against reciprocity or indirect utility driving the nature of feedback (Proposition 1A). The standard deviation of ratings is larger for free e-books than for non-free e-books ($p\text{-value} < 0.001$), supporting the proposition that the dispersion of ratings is larger for free e-books (Proposition 2). As these effects might be driven by authors of lower-quality e-books being more likely to offer their e-books for free, in the middle panel of table 8, I use only ratings of e-books where ratings were given to both free and non-free versions. In this sub-sample, I find that non-free e-books have a lower average rating than free e-books, although the difference is only significant at the 0.1 significance level. The standard deviation of ratings for free e-books is, however, still larger than for non-free e-books ($p\text{-value} < 0.001$). To control more carefully for selection effects and other confounding factors, in a next step, I demean each rating by subtracting the average rating of the e-book the rating was assigned to. The lower panel of table 8 shows the averages of demeaned ratings for both groups. The mean for non-free e-books is higher ($p\text{-value} < 0.001$), thus further supporting proposition 1 and providing evidence against proposition 1A. Again, the standard deviation of demeaned ratings for free e-books is larger than for non-free e-books ($p\text{-value} < 0.001$), providing support for proposition 2.

To sum up, I already find evidence for proposition 1 and 2, looking only at descriptive statistics.

4.2 Regression Analysis

Although the results from the descriptive analysis provide first evidence in favor of propositions 1 and 2, the results might be driven by other differences between free and non-free e-books. Therefore, in this section, I provide more detailed evidence by presenting the results of a series of regression models. Although I also show results from cross-sectional regressions, my preferred regression models include e-book level fixed effects, i.e.

$$Y_{ij} = \beta \text{FreeEbookRated}_j + f_i + \mathbf{X}_j\gamma + \epsilon_{ij},$$

where Y_{ij} is an outcome variable (e.g. a rating) on the level of e-book i and rating j , the dummy FreeEbookRated_j denotes whether a rating was assigned to a free e-book, f_i are e-book level fixed effects, \mathbf{X}_j is a vector of controls on the level of a rating, and ϵ_{ij} is an error term. The coefficient β and the vector of coefficients γ are only identified through within e-book variation, taking advantage of the fact that for a subset of e-books ratings were assigned to both free and non-free versions of the same e-book.

I analyze three separate outcome variables on the rating level for Y_{ij} : Rating, i.e. how many stars a reader assigned to an e-book, the deviation of a rating from the average rating of an e-book as a measure of dispersion of ratings, and the length of a review in terms of the number of characters it contains.

Rating

Table 9 shows results from a regression analysis with an individual rating as the unit of observation. The dependent variable in these regressions is the number of stars (1 to 5) a reviewer assigned to an e-book. In columns (1) and (2), results from cross-sectional OLS regressions are shown, including all e-book and rating level covariates. In columns (3) and (4), results from OLS regressions including e-book level fixed effects are shown. Only rating level covariates are included in these regressions, as the fixed effects pick-up any covariates that are constant on the e-book level.

Table 9: Impact of Offering Free E-book on Ratings

	Dependent variable:			
	Rating (1-5 Stars)			
	OLS	OLS with E-book Level Fixed Effects		
	(1)	(2)	(3)	(4)
Free E-book Rated (Yes=1)	−0.163*** (0.005)	−0.074*** (0.007)	−0.074*** (0.019)	−0.090*** (0.021)
Log(1+ Price) (Demeaned)		0.048*** (0.005)		
Log(Length E-book in Words) (Demeaned)		0.027*** (0.003)		
Gender Reviewer (Female=1)		0.014* (0.008)		0.041*** (0.012)
Free E-book Rated X Log(1 + Price) (Demeaned)		−0.017* (0.010)		−0.008 (0.027)
Free E-book Rated X Log(Length E-book) (Demeaned)		0.029*** (0.004)		0.022 (0.016)
Free E-book Rated X Gender Reviewer		0.144*** (0.013)		0.107*** (0.019)
Constant	4.813*** (0.095)	4.856*** (0.095)		
Category (Categorical)	Yes	Yes	No	No
Subcategory (Categorical)	Yes	Yes	No	No
Language (Categorical)	Yes	Yes	No	No
Year Published (Categorical)	Yes	Yes	No	No
N Ratings	Yes	Yes	No	No
Time Since Published (Weeks)	Yes	Yes	No	No
Average Rating Previous E-books	Yes	Yes	No	No
Gender Author	Yes	Yes	No	No
N E-books	Yes	Yes	No	No
N Previous E-books	Yes	Yes	No	No
Reviewed After Published (Weeks)	Yes	Yes	Yes	Yes
N Ratings Previous E-books	Yes	Yes	Yes	Yes
Observations	109,421	109,421	109,421	109,421
R ²	0.099	0.107	0.612	0.613

*p<0.1; **p<0.05; ***p<0.01
Robust standard errors in parentheses

Note: This table shows OLS regression results with a single rating as the unit of observation. The dependent variable is the number of “stars” that a reviewer assigned to an e-book. Reviewers can choose within a range of one (worst rating) up to five (best rating) stars. In column (3) and (4), e-book level fixed effects are included.

In the cross-sectional regression model in column (1), additional to the controls, I include the dummy *FreeEbookRated* indicating whether a rating was assigned to a free e-book. The estimated coefficient indicates that on average the rating for a free e-book is -0.163 stars lower than the rating assigned to a non-free e-book ($p\text{-value} < 0.01$).

In column (2), I include additional controls on the e-book and rating level. The variable *Price* measures the price of an e-book at the time when I collected the data. I also include the variable *Length E-book in Words*. Both variables, *Price* and *Length E-book in Words*, should capture to some degree how valuable an e-book is to a reader. These variables are interesting to look at, as the negative selection effect should be stronger, the larger the difference in price between free and non-free versions of the same e-book. On the other hand, if reciprocity plays a role in readers' rating behavior, receiving a "larger gift" in the form of a normally more expensive or longer e-book should trigger a stronger reciprocal response by reviewers in the form of a better rating. To check for the presence of both mechanisms, I include in column (2) interactions between *Price* and *Length E-book in Words*, and whether a rating was given to a free e-book. The coefficient on the interaction between *Price* and *FreeEbookRated* is indeed negative, consistent with a stronger negative selection effect for more expensive e-books. However, the effect is only significant at the 10 percent level. The coefficient on the interaction between *Length E-book in Words* and *FreeEbookRated*, on the other hand, is negative ($p\text{-value} < 0.01$), indicating that readers reciprocate more by giving better ratings if they receive a longer e-book for free.

In the regression in column (2), I also include the gender of a reviewer and an interaction between the gender and whether a rating was assigned to a free e-book. Interestingly, the coefficient on the interaction is positive ($p\text{-value} < 0.01$). An explanation might be that women act more reciprocal than men, consistent with a small literature in Behavioral Economics looking at gender effects in trust games (Croson and Buchan, 1999).

In column (3), I estimate a regression model where I include fixed effects on the e-book level. These fixed effects capture the mean rating of each e-book. The coefficient

on the variable *FreeEbookRated* is, therefore, only identified through e-books where it is not constant on the e-book level. The coefficient has a similar size as in the cross-sectional regressions and remains statistically significant ($p\text{-value} < 0.01$), indicating that free e-books receive ratings that are on average 0.074 stars lower than ratings for non-free versions of the same e-book. This is further support for proposition 1 and against proposition 1A.

In the regression in column (4), I include the same interactions between the variable *FreeEbookRated* and the variables *Price*, *Length E-book in Words*, and *Gender Reviewer*. The estimated coefficients have the same signs as in the cross-sectional regression in column (2). However, only the coefficient on the interaction between *FreeEbookRated* and *Gender Reviewer* remains statistically significant ($p\text{-value} < 0.05$).

To sum up, in all regression models I find evidence in favor of proposition 1, implying that free e-books receive worse ratings. The effect sizes show that free e-books receive ratings that are between 0.07 and 0.09 stars lower than ratings for non-free e-books, or 8 to 10 percent lower if expressed in terms of the standard deviation of ratings. However, I also find some evidence in favor of proposition 1A by finding that female reviewers, who according to previous literature should act more reciprocal than men, give better ratings to free e-books than male reviewers.

Dispersion of Ratings

In this subsection, I examine whether the prediction of proposition 2 that the dispersion of ratings is larger for free e-books than for non-free e-books holds empirically. To test it within a regression framework, I use the following approach: After demeaning each rating, I use the absolute value of each demeaned rating as the dependent variable in my regression models. This absolute value of demeaned ratings measures the absolute deviation of any given rating from the average rating of an e-book.⁷ Using this variable as a dependent variable in my regressions lets me estimate the effect that the variable

⁷For example, if an e-book has received three ratings given by 1, 3, 5; the average rating is 3 and the demeaned ratings are $-2, 0, 2$. The absolute value of these demeaned ratings is $2, 0, 2$.

Table 10: Impact of Offering Free E-book on Dispersion of Ratings

	Dependent variable:			
	Abs (Demeaned Rating)			
	OLS	OLS with E-book Level Fixed Effects		
	(1)	(2)	(3)	(4)
Free E-book Rated (Yes=1)	0.100*** (0.003)	0.074*** (0.004)	0.043*** (0.011)	0.041*** (0.012)
Log(1 + Price) (Demeaned)		-0.014*** (0.003)		
Log(Length E-book in Words) (Demeaned)		0.0004 (0.001)		
Gender Reviewer (Female=1)		0.010** (0.004)		-0.022*** (0.006)
Free E-book Rated X Log(1 + Price) (Demeaned)		-0.003 (0.005)		-0.026* (0.015)
Free E-book Rated X Log(Length E-book) (Demeaned)		-0.017*** (0.002)		-0.003 (0.009)
Free E-book Rated X Gender Reviewer		-0.048*** (0.007)		-0.014 (0.010)
Constant	-0.141*** (0.049)	-0.152*** (0.048)		
Category (Categorical)	Yes	Yes	No	No
Subcategory (Categorical)	Yes	Yes	No	No
Language (Categorical)	Yes	Yes	No	No
Year Published (Categorical)	Yes	Yes	No	No
N Ratings	Yes	Yes	No	No
Time Since Published (Weeks)	Yes	Yes	No	No
Average Rating Previous E-books	Yes	Yes	No	No
Gender Author	Yes	Yes	No	No
N E-books	Yes	Yes	No	No
N Previous E-books	Yes	Yes	No	No
Reviewed After Published (Weeks)	Yes	Yes	Yes	Yes
N Ratings Previous E-books	Yes	Yes	Yes	Yes
Observations	109,421	109,421	109,421	109,421
R ²	0.183	0.185	0.638	0.639

*p<0.1; **p<0.05; ***p<0.01
Robust standard errors in parentheses

Note: This table shows OLS regression results with a single rating as the unit of observation. The dependent variable is the absolute value of the deviation of a given rating from the average rating of an e-book. E-book level fixed effects are included in all regressions.

FreeEbookRated has on the deviation of each individual rating from the average rating of an e-book. Although, as before, I also show results from cross-sectional regressions, my preferred regression specification again includes e-book level fixed effects, i.e.

$$|Y_{ij} - \bar{Y}_i| = \beta \text{FreeEbookRated}_j + f_i + \mathbf{X}_j\gamma + \epsilon_{ij},$$

where Y_{ij} is rating j assigned to e-book i , and \bar{Y}_i is the mean rating of e-book i . While the e-book level fixed effect f_i captures the average standard deviation of ratings on the e-book level, the coefficient β on the variable *FreeEbookRated* picks up any difference in standard deviation for free e-books.

Table 10 shows the results of these regressions. Columns (1) and (2) show results from cross-sectional regressions, including all e-book level and rating level covariates. In column (1), I only include the variable *FreeEbookRated*, while in column (2) I include the interactions between *Price*, *Length E-book in Words*, and *Gender Reviewer*. In both cross-sectional regressions, the coefficient on *FreeEbookRated* is positive (p -values < 0.01), indicating that the standard deviation of ratings is higher for free than for non-free e-books. The fixed effects regressions in column (3) and (4) also support this conclusion, as the coefficients in column (3), only including the variable *FreeEbookRated*, and column (4), including the same interactions as in column (2), are positive (p -values < 0.01). Using the fact that the overall standard deviation of ratings is 0.86, the effect size indicates that the standard deviation of ratings is 5 to 12 percent higher for free than for non-free e-books.

Thus overall, I also find support for proposition 2 within a regression framework.

Length of Review

In this subsection, I explore whether reviewers change the nature of their feedback in terms of the length of text they post together with a rating. The text of a review contains additional information that future buyers can use to base their purchasing decision on. Longer reviews should, on average, be more informative and thus help authors to find the audience that best fits their personal style, increasing the horizontal

match between authors and readers.

Table 11 shows results from regressions where the dependent variable is the length of a given review measured by the number of characters it contains. As the distribution of this variable is skewed, I transform the variable using the natural logarithm.

In columns (1) and (2), I present results of cross-sectional regressions, while in column (3) and (4) I include e-book level fixed effects into my regressions. In columns (1) and (3), I only include the variable *FreeEbookRated*. In columns (2) and (4), I also include the interactions between *Price*, *Length E-book in Words*, and *Gender Reviewer*.

The coefficient on the variable *FreeEbookRated* is negative and statistically significant in all four regression models, implying that on average free e-books receive shorter reviews than non-free e-books. The size of the coefficients in the fixed effects regressions, which should be the most accurate estimates of the effect size, imply that free e-books receive reviews that are approximately 6 to 8 percent shorter than the reviews of non-free e-books.

From the coefficients on the interaction terms, both the coefficient on the interaction between *FreeEbookRated* and *Price*, and the interaction between *FreeEbookRated* and *Gender Reviewer* are statistically significant, both in the cross-sectional and in the fixed effects regressions. The positive coefficient on the interaction between *FreeEbookRated* and *Price* implies that e-books that are more expensive when they are not free receive longer reviews when they are offered for free. This result might imply that reviewers reciprocate more by writing longer reviews when normally they would have paid a high price for an e-book. However, as implied by the negative coefficient on the interaction between *FreeEbookRated* and *Gender Reviewer*, female reviewers write on average shorter reviews when receiving an e-book for free, which is not consistent with women acting more reciprocal.

To sum up, I find evidence that reviewers who receive a free e-book do not only give a worse rating but they also provide shorter reviews. However, the more expensive the e-book is otherwise, the weaker this effect becomes, consistent with a reciprocity effect that becomes stronger as reviewers receive larger “gifts.” While in the case of

Table 11: Impact of Offering Free E-book on Length of Reviews

	<i>Dependent variable:</i>			
	Log(1 + Characters in Review)			
	OLS	OLS with E-book Level Fixed Effects		
	(1)	(2)	(3)	(4)
Free E-book Rated (Yes=1)	−0.684*** (0.007)	−0.450*** (0.011)	−0.080*** (0.026)	−0.062** (0.027)
Log(1+ Price) (Demeaned)		−0.006 (0.008)		
Log(Length E-book in Words) (Demeaned)		0.155*** (0.004)		
Gender Reviewer (Female=1)		0.070*** (0.012)		−0.019 (0.017)
Free E-book Rated X Log(1 + Price) (Demeaned)		0.157*** (0.016)		0.215*** (0.037)
Free E-book Rated X Log(Length E-book) (Demeaned)		−0.009* (0.005)		0.017 (0.020)
Free E-book Rated X Gender Reviewer		−0.207*** (0.018)		−0.051** (0.025)
Constant	5.599*** (0.134)	5.692*** (0.133)		
Category (Categorical)	Yes	Yes	No	No
Subcategory (Categorical)	Yes	Yes	No	No
Language (Categorical)	Yes	Yes	No	No
Year Published (Categorical)	Yes	Yes	No	No
N Ratings	Yes	Yes	No	No
Time Since Published (Weeks)	Yes	Yes	No	No
Average Rating Previous E-books	Yes	Yes	No	No
Gender Author	Yes	Yes	No	No
N E-books	Yes	Yes	No	No
N Previous E-books	Yes	Yes	No	No
Reviewed After Published (Weeks)	Yes	Yes	Yes	Yes
N Ratings Previous E-books	Yes	Yes	Yes	Yes
Observations	109,421	109,421	109,421	109,421
R ²	0.154	0.177	0.664	0.665

*p<0.1; **p<0.05; ***p<0.01
Robust standard errors in parentheses

Note: This table shows OLS regression results with a single rating as the unit of observation. The dependent variable is the logarithm of the number of characters reviewers use in their reviews. In column (3) and (4), e-book level fixed effects are included.

ratings as an outcome variable I find results that are consistent with female reviewers reciprocating more than male reviewers, I do not find such an effect in the case of the length of reviews. On the contrary, female reviewers provide shorter reviews for free e-books. One explanation might be that female reviewers reciprocate by giving better ratings, while male reviewers reciprocate by giving longer reviews.

Differences between Genres

In this subsection, I explore differences in the effect size between different genres of e-books. Generally, it should be expected that the negative selection effect of offering

Table 12: Variance of Rating for Each Subgenre

Subgenre	Number E-books	Average Variance of Ratings on E-book Level
Religion and Spirituality	225	1.3818
Erotica	2,640	0.7352
Humor & Comedy	392	0.6180
Horror	1,075	0.6134
Science Fiction	1,526	0.6078
Young Adult or Teen	1,448	0.5607
Plays & Screenplays	255	0.5415
Mystery & Detective	803	0.5233
Romance	3,088	0.5166
(Other)	1,981	0.5132
Women's Fiction	242	0.5015
Gay & Lesbian Fiction	477	0.4748
Thriller & Suspense	726	0.4722
Literature	673	0.4632
Anthologies	332	0.4606
Fantasy	2,090	0.4582
Christian	210	0.4246
Adventure	312	0.4170
Historical	255	0.4100
Poetry	301	0.3384
Children's Books	265	0.2370
Biography	200	0.2266
Sum = 19,872		Average = 0.5522

Note: This table shows the average variance of ratings for each subgenre of e-books. Only e-books with more than one rating are included. The variance of ratings is first calculated on the level of each individual e-book and then averaged over all e-books within a subgenre.

an e-book for free is stronger for e-books for which readers have more heterogeneous tastes. Such a heterogeneity in tastes should be reflected in a higher dispersion of ratings, which should have an effect on the strength of the negative selection effect.⁸

As a measure of the heterogeneity of tastes for a particular e-book, I use the average variance of ratings within its subgenre. Table 12 shows the variance of ratings for each subgenre. The subgenres "Religion and Spirituality," "Erotica," and "Humor & Com-

⁸For example, Sun (2012) takes the variance of product ratings on *Amazon.com* as a measure of whether a product is a niche or a mass-market product. The underlying idea is that whereas for niche products some consumers have a very strong preference but others have a very low preference, for mass-market products consumers have more homogeneous preferences, which is reflected in a lower dispersion of ratings for mass-market products as compared to niche products.

Table 13: Impact of Genre Variance of Ratings

	Dependent variable:	
	Rating (1-5 Stars)	
	Fixed Effects on E-book Level	
	(1)	(2)
Free E-book Rated (Yes=1)	-0.078*** (0.020)	-0.156*** (0.025)
Free E-book Rated X Average Subgenre-Variance in Ratings (Demeaned)	-0.384** (0.165)	-0.342** (0.165)
Gender Reviewer (Female=1)		0.043*** (0.012)
Free E-book Rated X Log(1 + Price) (Demeaned)		-0.015 (0.028)
Free E-book Rated X Log(Length E-book in Words) (Demeaned)		0.016 (0.016)
Free E-book Rated X Gender Reviewer (Female=1)		0.106*** (0.019)
N Previous Ratings	Yes	Yes
Reviewed After Published (Weeks)	Yes	Yes
Observations	107,586	107,586
R ²	0.612	0.613

*p<0.1; **p<0.05; ***p<0.01
Robust standard errors in parentheses

Notes: This table shows regression results with an added interaction between the dummy indicating whether a rating was given for a free e-book with the average genre variance of ratings for each subgenre.

edy” have the highest variance of ratings, whereas the subgenres “Poetry,” “Children’s Books,” and “Biography” have the lowest variance.

In a next step, I match the corresponding variance of ratings for each subgenre to each individual rating based on the subgenre of the e-book. Then, I include an interaction of this variable with the dummy indicating whether a rating has been given for a free e-book into my regression models. Table 13 shows the results of two regression models with e-book level fixed effects. In column (1), only the variable whether a given rating has been given for a free e-book and its interaction with the variance on the subgenre level are included. In column (2), I include the same additional interactions as in my previous regressions. In both regression models, the coefficient on the interaction between *FreeEbookRated* and the subgenre variance of ratings is negative and statistically significant (p -values < 0.05). This implies that the negative selection effect indeed is stronger for subgenres with a higher variation in ratings, consistent with a stronger negative selection effect for e-books from subgenres for which readers

have more heterogeneous tastes.

4.3 Further Evidence in Support of Selection

In this section, I provide further evidence that the negative selection effect is driven by readers with a lower preference obtaining free e-books and, consequently, giving lower ratings. I use the fact that reviewers on *Smashwords* have to use a nickname when reviewing an e-book. Thereby, for reviewers who rate more than one e-book, I have some information on their reading habits in terms of which other e-books they have rated in the past. From this information, I calculate three additional variables as measures of a reviewer's reading habits and include them into my regression models: whether a given rating is assigned to an e-book that is from the same genre that the reviewer usually reviews (the reviewer's "favorite" genre), whether a given rating is assigned to an e-book that is from the same subgenre that the reviewer usually reviews (the reviewer's "favorite" subgenre),⁹ and whether a reviewer has reviewed an e-book by the same author in the past. These three variables should at least partly explain the negative selection effect.

Table 14 shows results of three logit regression models estimating whether ratings for free e-books are more likely to be given by reviewers who have other "favorite" genres or subgenres, or by reviewers that previously have not rated an e-book by the same author. The results in columns (2) and (3) show that a rating for a free e-book is 5 percent more likely to be assigned by a reviewer who has a different "favorite" subgenre (p -value < 0.01), and that a rating of a free e-book is 12 percent more likely to be assigned by a reader who has not previously rated another e-book by the same author (p -value < 0.01). However, as the result in column (1) shows, a rating for a free e-book is 0.1 percent less likely to be assigned by a reader who has a different "favorite" genre (p -value < 0.01). The effect size is, however, small compared to the effect sizes of the two other variables.

⁹Both the reviewer's "favorite" genre and the reviewer's "favorite" subgenre are calculated based on the genre or subgenre that the reviewer rated and reviewed most often. Ties are broken randomly.

Table 14: What Types of Readers Rate Free E-books

	<i>Dependent variable:</i>		
	Not Fav. Genre (Yes=1)	Not Fav. Subgenre (Yes=1)	Prev. Read Author (Yes=1)
	<i>Logit</i> (1)	<i>Logit</i> (2)	<i>Logit</i> (3)
Free E-book Rated (Yes=1)	−0.001*** (0.000)	0.054*** (0.005)	−0.124*** (0.005)
Constant	−0.003 (0.017)	−0.033 (0.062)	−0.061 (0.057)
Category (Categorical)	Yes	Yes	Yes
Subcategory (Categorical)	Yes	Yes	Yes
Language (Categorical)	Yes	Yes	Yes
Year Published (Categorical)	Yes	Yes	Yes
Price	Yes	Yes	Yes
Length E-book in Words	Yes	Yes	Yes
Gender Reviewer	Yes	Yes	Yes
N Ratings	Yes	Yes	Yes
Time Since Published (Weeks)	Yes	Yes	Yes
Gender Author	Yes	Yes	Yes
Average Rating Previous E-books	Yes	Yes	Yes
N E-books	Yes	Yes	Yes
N Previous E-books	Yes	Yes	Yes
Reviewed After Published (Weeks)	Yes	Yes	Yes
N Ratings Previous E-books	Yes	Yes	Yes
Observations	80,110	80,110	80,110

*p<0.1; **p<0.05; ***p<0.01
Coefficients give marginal effects estimates
Standard errors in parentheses

Note: This table shows logit regression showing what type of reviewer is more likely to rate a free e-book. The unit of observation is a given rating. Only ratings of reviewers who have reviewed more than one e-book are included. The regression in column (1) shows that if a rating is given for a free e-book, it is less likely that the genre of the e-book is not the readers “favorite” genre. The regression in column (2) shows that if a rating is given for a free e-book, it is more likely that the subgenre of the e-book is not the reviewers “favorite” subgenre. The regression in column (3) shows that if a rating is given for a free e-book, it is less likely that the reviewer has previously rated an e-book of the same author.

Table 15: Impact of Type of Reader on Rating

	<i>Dependent variable:</i>			
	Rating (1-5 Stars)			
	<i>OLS with E-book Level Fixed Effects</i>			
	(1)	(2)	(3)	(4)
Free E-book Rated (Yes = 1)	-0.072*** (0.026)	-0.072*** (0.026)	-0.069*** (0.026)	-0.053** (0.026)
Not Fav. Genre (Yes=1)		0.002 (0.044)	0.022 (0.044)	0.037 (0.043)
Not Fav. Subgenre (Yes=1)			-0.054*** (0.011)	-0.043*** (0.011)
Prev. Read Author (Yes=1)				0.181*** (0.011)
Free E-book Rated X Log(1 + Price) (Demeaned)	0.022 (0.020)	0.022 (0.020)	0.023 (0.020)	0.020 (0.020)
Free E-book Rated X Log(Length E-book in Words) (Demeaned)	0.072*** (0.017)	0.072*** (0.017)	0.073*** (0.017)	0.073*** (0.017)
Free E-book Rated X Gender Reviewer (Female=1)	0.083*** (0.025)	0.083*** (0.025)	0.083*** (0.025)	0.084*** (0.025)
N Previous Ratings	Yes	Yes	Yes	Yes
Reviewed After Published (Weeks)	Yes	Yes	Yes	Yes
Observations	80,110	80,110	80,110	80,110
R ²	0.657	0.657	0.657	0.660

*p<0.1; **p<0.05; ***p<0.01
Robust standard errors in parentheses

Note: This table shows OLS regressions where I include the characteristics of the reader giving a rating. The dependent variable is rating with a single rating as the unit of observation. In column (2), I add whether the e-book for which a reviewer gives a rating is not from his “favorite” category as a covariate. In column (3), I add whether the e-book for which a reviewer gives a rating is not from his “favorite” subcategory. In column (3), I add whether the reviewer has previously rated any e-book by the same author. Only reviews by reviewers with more than one review are included.

Table 15 shows regression results where I include the three measures of a reviewer's reading habits consecutively into fixed effects regression models with a rating as the dependent variable. Only ratings by reviewers who have reviewed more than one e-book are included. In column (1), I include the measure indicating whether a rating was assigned by a reviewer with a different "favorite" genre. This variable does not have a significant effect on the rating that the reviewer assigns. In column (2), I include the measure indicating whether a rating was assigned by a reviewer with a different "favorite" subgenre. This variable has a negative impact on the rating a reviewer assigns, indicating that such a reviewer gives on average a rating that is 0.054 stars lower (p -value < 0.01). Including this variable decreases the effect size of the main effect of whether a rating is assigned to a free e-book, although the decrease is small and not statistically significant. In column (3), I include the measure indicating whether the reviewer has previously reviewed an e-book by the same author. This variable has a strong effect on the rating a reviewer assigns, indicating that a reviewer assigns a rating that is by 0.181 stars higher if he or she has previously rated an e-book by the same author (p -value < 0.01). Including this variable also reduces the effect size of whether the rating was assigned to a free e-book by 16 percent, although the decrease is not statistically significant.

To sum up, I find strong evidence in favor of reviewers who usually buy different kinds of e-books give worse ratings to free e-books. Including these explanatory variables reduces the main effect by about 20 percent, although the decrease is not statistically significant using the usual levels of statistical significance.

5 Conclusion

Overall, I find strong empirical support for the hypothesis that giving away products for free induces a negative selection effect on sellers' (online) reputation. The alternative hypotheses that giving away products for free induces buyers to reciprocate by giving a better rating or that buyers assign a rating according to their indirect utility are overall rejected. However, I find some evidence for reciprocity in the form of fe-

male authors giving better ratings to free products than male reviewers, which would be consistent with research in Behavioral Economics showing that women act more reciprocal than men. Further, I find evidence that the dispersion of ratings is higher for products that are given away for free.

This has the implication that giving away products for free is a double-edged strategy for sellers. While it still might be possible to build a reputation by giving away products for free, it involves a trade-off, especially for products where consumers have heterogeneous tastes.

My study has two important implications. One is that sellers trying to build a reputation should specifically target free units of their products to buyers who are likely to have a high preference for their products. For example, in my empirical setting, authors should target free units of their e-books to readers that have bought previous e-books of the same author or books from the same genre. This would not only decrease the negative selection effect but authors might even be able to induce a positive selection effect, enabling them to build a better reputation.

In terms of policy, a negative selection effect on reputation implies an additional entry barrier for sellers and firms entering markets with asymmetric information. Similar forms of reputation as an entry barrier have already been suggested by previous literature (e.g. Farrell, 1986; Grossman and Horn, 1988), although not in the form of a selection effect as in this study. Policy makers such as governments or platforms could help to lower such entry barriers by either providing quality certification themselves or by debiasing reputation measures, for example by giving weight to more “objective” reviewers who have a weaker personal taste for the products they rate.

Bibliography

- Ba, S., and Pavlou, P. A. (2002). "Evidence of the Effect of Trust Building Technology in Electronic Markets: Price Premiums and Buyer Behavior." *MIS Quarterly*, 243–268.
- Bar-Isaac, H. (2004). *Reputation and Professional Services: Survival, Teams and Incentives*. Ph.D. thesis, University of London.
- Bar-Isaac, H., and Deb, J. (2014a). "(Good and Bad) Reputation for a Servant of Two Masters." *American Economic Journal: Microeconomics*, 6(4), 293–325.
- Bar-Isaac, H., and Deb, J. (2014b). "What is a Good Reputation? Career Concerns with Heterogeneous Audiences." *International Journal of Industrial Organization*, 34, 44–50.
- Bar-Isaac, H., and Tadelis, S. (2008). "Seller Reputation." In *Foundations and Trends in Microeconomics*, vol. 4, 273–351, Boston: Now Publishers.
- Bawa, K., and Shoemaker, R. (2004). "The Effects of Free Sample Promotions on Incremental Brand Sales." *Marketing Science*, 23(3), 345–363.
- Bergemann, D., and Välimäki, J. (2000). "Experimentation in Markets." *Review of Economic Studies*, 67(2), 213–234.
- Bergemann, D., and Välimäki, J. (2006). "Dynamic Pricing of New Experience Goods." *Journal of Political Economy*, 114(4), 713–743.
- Bolton, G., Greiner, B., and Ockenfels, A. (2013). "Engineering Trust: Reciprocity in the Production of Reputation Information." *Management Science*, 59(2), 265–285.
- Bose, S., Orosel, G., Ottaviani, M., and Vesterlund, L. (2006). "Dynamic Monopoly Pricing and Herding." *RAND Journal of Economics*, 37(4), 910–928.
- Bowker (2014). "Self-Publishing in the United States, 2007-2013." <http://media.bowker.com/documents/bowker/selfpublishing/report2013.pdf>, Accessed on July 12, 2016.
- Cabral, L. (2005). "The Economics of Trust and Reputation: A Primer." Draft Available at <http://luiscabral.net>.
- Cabral, L. (2012). "Oligopoly Dynamics." *International Journal of Industrial Organization*, 30(3), 278–282.
- Cabral, L., and Hortacsu, A. (2010). "The Dynamics of Seller Reputation: Evidence from eBay." *Journal of Industrial Economics*, 58(1), 54–78.
- Cabral, L., and Li, L. (2015). "A Dollar for Your Thoughts: Feedback-Conditional Rebates on eBay." *Management Science*, 61(9), 2052–2063.
- Campbell, A., Mayzlin, D., and Shin, J. (2013). "Managing Buzz." Working Paper, Available <https://www.marshall.usc.edu/faculty/directory/mayzlin>.
- Chevalier, J., and Mayzlin, D. (2006). "The Effect of Word of Mouth on Sales: Online Book Reviews." *Journal of Marketing Research*, 43(3), 345–354.

- Claussen, J., Kretschmer, T., and Mayrhofer, P. (2013). "The Effects of Rewarding User Engagement: The Case of Facebook Apps." *Information Systems Research*, 24(1), 186–200.
- Coker, M. (2012). "How to Sell E-books at the Apple iBookstore." *Smashwords Blog*, Available at <http://blog.smashwords.com/2012/11/how-to-sell-ebooks-at-apple-ibookstore.html>, Accessed on January 16, 2014.
- Cripps, M. W. (2009). *Reputation*. New Palgrave Dictionary of Economics, Basingstoke: Palgrave Macmillan, 2nd edn.
- Croson, R., and Buchan, N. (1999). "Gender and Culture: International Experimental Evidence from Trust Games." *American Economic Review*, 89(2), 386–391.
- Dellarocas, C. (2003). "The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms." *Management Science*, 49(10), 1407–1424.
- Dellarocas, C. (2006). "Strategic Manipulation of Internet Opinion Forums: Implications for Consumers and Firms." *Management Science*, 52(10), 1577–1593.
- Dewan, S., and Hsu, V. (2004). "Adverse Selection in Electronic Markets: Evidence from Online Stamp Auctions." *Journal of Industrial Economics*, 52(4), 497–516.
- Falk, A., and Fischbacher, U. (2006). "A Theory of Reciprocity." *Games and Economic Behavior*, 54(2), 293–315.
- Farajallah, M., Hammond, R., and Pénard, T. (2016). "What Drives Pricing Behavior in Peer-To-Peer Markets? Evidence from the CarSharing Platform BlaBlaCar." *Working Paper*.
- Farrell, J. (1986). "Moral Hazard as an Entry Barrier." *RAND Journal of Economics*, 440–449.
- Friedman, E., Resnick, P., and Sami, R. (2007). "Manipulation-Resistant Reputation Systems." *Algorithmic Game Theory*, 677–697.
- Grossman, G. M., and Horn, H. (1988). "Infant-industry protection reconsidered: The case of informational barriers to entry." *Quarterly Journal of Economics*, 103(4), 767–787.
- Kahneman, D., Knetsch, J. L., and Thaler, R. (1986). "Fairness as a Constraint on Profit Seeking: Entitlements in the Market." *American Economic Review*, 76(4), 728–741.
- Klein, T. J., Lambertz, C., and Stahl, K. O. (forthcoming). "Market Transparency, Adverse Selection, and Moral Hazard." *Journal of Political Economy*.
- Li, L., and Xiao, E. (2014). "Money Talks: Rebate Mechanisms in Reputation System Design." *Management Science*, 60(8), 2054–2072.
- MacLeod, W. B. (2007). "Reputations, Relationships, and Contract Enforcement." *Journal of Economic Literature*, 595–628.

- Mailath, G. J., and Samuelson, L. (2006). *Repeated Games and Reputations*. Oxford: Oxford University Press.
- McFadden, D. L., and Train, K. E. (1996). "Consumers' Evaluation of New Products: Learning from Self and Others." *Journal of Political Economy*, 104(4), 683–703.
- Mullen, L. (2015). "gender: Predict Gender from Names Using Historical Data."
- Nelson, P. (1970). "Information and Consumer Behavior." *Journal of Political Economy*, 78(2), 311–329.
- Nosko, C., and Tadelis, S. (2015). "The Limits of Reputation in Platform Markets: An Empirical Analysis and Field Experiment." *NBER Working Paper No. 20830*.
- Peitz, M., and Waelbroeck, P. (2006). "Why the Music Industry May Gain from Free Downloading — The Role of Sampling." *International Journal of Industrial Organization*, 24(5), 907–913.
- Rabin, M. (1993). "Incorporating Fairness into Game Theory and Economics." *American Economic Review*, 83(5), 1281–1302.
- Resnick, P., and Zeckhauser, R. (2002). "Trust Among Strangers in Internet Transactions: Empirical Analysis of Ebay's Reputation System." *Economics of the Internet and E-commerce*, 11(2), 23–25.
- Shapiro, C. (1983). "Premiums for High Quality Products as Returns to Reputations." *Quarterly Journal of Economics*, 98(4), 659–679.
- Sun, M. (2012). "How Does the Variance of Product Ratings Matter?" *Management Science*, 58(4), 696–707.
- Villas-Boas, J. M. (2004). "Consumer Learning, Brand Loyalty, and Competition." *Marketing Science*, 23(1), 134–145.
- Waldfoegel, J., and Reimers, I. (2015). "Storming the Gatekeepers: Digital Disintermediation in the Market for Books." *Information Economics and Policy*, 31, 47–58.
- Wibral, M. (2015). "Identity Changes and the Efficiency of Reputation Systems." *Experimental Economics*, 18(3), 408–431.
- Zegners, D. (2016). "Voluntary Disclosure of Product Information: The Case of E-Book Samples." *Working Paper Available at SSRN*.