Voluntary Disclosure of Product Information: The Case of E-book Samples*

Dainis Zēgners

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

June 7, 2016

Job Market Paper

Current Version:
http://dainiszegners.net/ebooks

Abstract

An important question in markets with asymmetric information is why in practice fewer sellers voluntarily disclose their private information than theory would predict. To better understand this discrepancy, I use data from an online self-publishing platform to examine the empirical relationship between pricing and voluntary disclosure. On this platform, I observe whether authors disclose characteristics of their e-books by offering free samples. In contrast to the prediction of theories of unraveling, I show that for e-books without a posted online rating, indicating that their quality is unknown to the market, offering a sample is associated with a lower price. I also show that for unrated e-books, fewer authors offer a sample while simultaneously setting a higher price than authors of rated e-books. These results can be explained by incorporating into a model a fraction of naive buyers who do not update their beliefs upon observing that a seller does not disclose. This gives low-quality sellers an incentive to conceal their quality by not disclosing and to set high prices to exploit naive buyers.

Keywords: Voluntary Disclosure, Asymmetric Information, Consumer Naivete
JEL codes: L15, L86, M37

*I thank my colleagues in Munich at the ISTO, my Ph.D. advisors Tobias Kretschmer and Florian Englmaier, José-Manuel Giménez-Gómez, Michał Krawczyk, Joaqíun Coleff, Lidia Tseitlin, Thorsten Grohsjean, Forrest Spence, Johannes Münster, Jeanne Hagenbach, audiences at Telecom ParisTech, the 2nd PhD workshop on Industrial and Public Economics in Reus, the 5th Economics of ICTs Workshop in Porto, the EEA 2014 in Toulouse, the EARIE 2014 in Milan, the 2nd Workshop on Industrial Organization in the Digital Economy in Louvain-La-Neuve, the Workshop on the Economics of Network Industries in Paris, the Natural Experiments and Controlled Field Studies Workshop in Ohlstadt, the ICT Conference at ZEW Mannheim, and the ORG and TIME Seminars at LMU Munich 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.

Readers are not sheep, and not every pen tempts them.

VLADIMIR NABOKOV[1]

1 Introduction

A fundamental idea in information economics is that for markets to work efficiently, buyers and sellers need to possess symmetric information. If not, the market will not allocate resources efficiently or problems such as adverse selection and moral hazard will arise. An important question is therefore whether sellers will voluntarily disclose private information about their products or whether mandatory disclosure policies are necessary. One answer to this question is given by the famous unraveling principle (Grossman, 1981; Milgrom, 1981; Jovanovic, 1982; Viscusi, 1978): As the non-disclosing seller with the highest quality should always disclose his quality, in equilibrium the market will fully unravel such that all sellers disclose. The price mechanism plays a crucial role in this argument: Without revealing their true type, high-quality sellers cannot distinguish themselves from low-quality sellers. As a consequence, they are pooled by the market and earn a price equal to the average quality of all non-disclosing sellers. To avoid such punishment by the market, sellers with the highest quality among non-disclosing sellers have an incentive to disclose their quality. If such punishment by pooling does not occur, the market will fail to unravel and a mandatory disclosure policy may be necessary. Whether such a policy improves consumer welfare also depends on how sellers' prices respond to disclosure. For these reasons, it is crucial to understand the link between disclosure and pricing in markets with asymmetric information.

In this paper, I study the empirical relationship between voluntary disclosure and

[1]The quote is from a letter Nabokov wrote in 1963 to his editor, William McGuire. The full quote contains some additional wisdom about the publishing industry: "I do not believe that a distinguished critic's review (or indeed any review) helps to sell a book. Readers are not sheep, and not every pen (pun) tempts them. Some of my best flops had been ushered in by extravagant (albeit well deserved) praise from eminent critics. The only thing that is of some help to commercial success of a book (apart from topicality or sexuality) is a sustained advertising campaign, lots of ads everywhere" (Nabokov, 2012).

pricing in a setting where I observe whether independent authors on an internet self-publishing platform offer free samples of their e-books. My data offers two advantages: First, I observe how much information on an e-book's quality in the form of online ratings posted by previous buyers is available to buyers. Second, reading a sample does not only inform buyers about an e-book's quality but also about its horizontal characteristics. Therefore, I can test a richer framework put forward in a recent theoretical literature by Johnson and Myatt (2006); Bar-Isaac et al. (2010); Sun (2011); Koessler and Renault (2012), and Celik (2014).[2] This literature studies models where sellers cannot only disclose information about their products' quality but also about their products' horizontal characteristics and their match to buyers' idiosyncratic tastes. As information about horizontal characteristics by its very nature increases the valuation of some but decreases the valuation of other buyers, a seller will only disclose horizontal information if she can thereby profitably target buyers with a higher valuation for the product's horizontal characteristics. From these buyers, the seller earns a "match-premium" in the form of a higher price. This also introduces a positive link between disclosure and a product's price, similar to the unraveling principle.

My results show that offering a sample is associated with a higher price only for e-books that have been rated by previous buyers. This result confirms the theoretical prediction of sellers with known quality earning a match-premium when disclosing their horizontal characteristics. For e-books that have not been rated by previous buyers, I observe the opposite: Offering a sample is associated with a lower price. This result is the reverse of the punishment by pooling of non-disclosing sellers required for the unraveling principle to work. The absence of pooling is further supported by a higher price dispersion for e-books not offering a sample than for e-books offering a sample. I also find evidence that authors take advantage of the absence of pooling by not disclosing their quality and at the same time elevating their price, as authors of non-rated e-books are less likely to offer samples and their average prices are higher.

[2] A related theoretical literature looks also at disclosure of horizontal information in competitive markets (Hotz and Xiao, 2013; Janssen and Teteryatnikova, 2015), whereas the above literature studies models with a monopolistic seller.

I obtain similar results when using alternative measures for the availability of information on e-books such as whether a given author releases his first e-book or whether ratings for previous e-books are available. Instrumenting for the availability of ratings using the number of weeks an e-book has been on sale also produces similar results.

To show how my results can be reconciled with existing models, I extend a simple model of voluntary disclosure by including a fraction of naive consumers. Evidence ranging from industries such as hospitals (Jin, 2005), restaurants (Bederson et al., 2015), salad dressings (Mathios, 2000) and business schools (Luca and Smith, 2015) shows that full unraveling is rarely observed in practice. A commonly suggested explanation[3] is that consumers fail to anticipate that low-quality sellers are less likely to disclose their quality (Dranove and Jin, 2010, p.943), a deviation from the rational consumer paradigm that has also been found in the lab (Jin, 2005). Including a fraction of naive consumers who fail to account for sellers' equilibrium behavior in the spirit of Esponda (2008) and Eyster and Rabin (2005), I show how a positive relationship between disclosure and price can be explained within the model: As naive consumers have an expectation of non-disclosing sellers' quality that is too high, low-quality sellers exploit naive buyers by not disclosing their quality and setting a high price.[4] High-quality sellers, on the other hand, disclose their quality to cater to both naive and rational buyers. When naive buyers inflated expectation of sellers' quality is higher than the average quality of disclosing high-quality sellers, a negative relationship between disclosure and price will be observed. As low-quality sellers can only exploit naive consumers when consumers are uninformed about sellers' quality, such a model can also explain why I observe overall lower disclosure rates for unrated e-books.

My paper makes the following contributions to the literature: First, I confirm the prediction that disclosing sellers earn a match-premium based on disclosing their hori-

[3]Other common explanations are disclosure costs, strategic interaction between sellers, and lack of credibility of disclosed information (Dranove and Jin, 2010, pp.943). All three explanations are unlikely to be present in my setting. See Dranove and Jin (2010) for a detailed discussion of the literature on disclosure.

[4]Lab experiments have also shown that sellers take advantage of naive or uninformed buyers. For example, Jin et al. (2014) find that sellers disclose less when buyers have inflated beliefs about non-disclosing sellers. Henze et al. (2015) find that sellers can sustain prices above their quality if buyers lack information on sellers' quality.

zontal characteristics that is made by recent models of disclosure of horizontal product characteristics (Johnson and Myatt, 2006; Bar-Isaac et al., 2010; Sun, 2011; Koessler and Renault, 2012; Celik, 2014) . Second, I document a deviation with field data from the pricing mechanism required for the classical unraveling theories (Grossman, 1981; Milgrom, 1981; Jovanovic, 1982; Viscusi, 1978) to work, an anomaly that can explain the general failure of these theories in practice. Third, by showing how my results can be explained by incorporating naive consumers into these models, my paper is one of the first to formally link the literature on voluntary disclosure to models of adverse selection with boundedly rational agents (Eyster and Rabin, 2005; Esponda, 2008), thus contributing to the emerging literature on Behavioral Industrial Organization.[5]

More broadly, voluntary disclosure in the form of offering samples can also be understood as informative advertising (Stigler, 1961; Ozga, 1960). The empirical literature on the link between advertising and pricing has found mixed results in both directions (see e.g. Bagwell, 2007 or Genesove and Simhon, 2015 for an overview), which most likely is due to the multitude of economic functions that advertising can have, additional to the informative view. My study can be seen as a cleaner test of the relationship between informative advertising[6] and pricing. A paper with a finding similar to my result of a positive relationship between price and disclosure that can be explained by a better match between buyers and sellers is Tadelis and Zettelmeyer (2015) in the context of used-car auctions.[7]

My paper is structured as follows: In section 2, I introduce a modeling framework based on the previous literature to derive testable propositions. In section 3, I describe my dataset, which I analyze in section 4. In section 5, I present some robustness checks confirming my initial analysis. In section 5, I show how my initial model needs to be

[5]A recent theoretical paper making the same link by using a similar framework of consumer naivete is Ispano and Schwardmann (2016). In contrast to my model, they focus on a competitive environment with vertically but not horizontally differentiated products. Another paper is Benndorf et al. (2015), which uses level-k reasoning by boundedly rational agents to explain incomplete unraveling in an experiment mimicking a labor market.

[6]Reading a sample informs buyers of a product's characteristics, although not about its existence.

[7]In their setting, however, the better match is achieved in the context of buyers better selecting themselves into competing auctions of vertically differentiated goods, not through horizontal match as in my setting.

adapted to account for my results. In section 6, I conclude.

2 Theoretical Framework

In this section, I present a short theoretical framework of voluntary disclosure. My goal is to capture the basic relationship between a seller's incentive to disclose product information and her pricing decision. To remain consistent with my empirical context, I label disclosing product information as "offering a sample" from which consumers can learn a product's characteristics

In my framework, products have both vertical and horizontal characteristics. The framework, therefore, captures both the seller's incentives to disclose her vertical characteristics, as in Grossman (1981) and Milgrom (1981); and her horizontal characteristics, as in Johnson and Myatt (2006), Bar-Isaac et al. (2010), Sun (2011), Koessler and Renault (2012) and Celik (2014). My framework is in particular close to the model in Sun (2011).

Consider a monopolistic market where a seller offers a product with both vertical and horizontal characteristics. Buyer i's utility for the product is

$$u_i = q + \epsilon_i,$$

where q is the product's quality and ϵ_i is the match between the buyer's idiosyncratic taste and the product's horizontal characteristic. There is a mass 1 of buyers whose match value ϵ_i is distributed uniformly on the interval $[-\bar{\epsilon}, \bar{\epsilon}]$ with density $\frac{1}{2\bar{\epsilon}}$.[8] Feasible qualities are distributed uniformly on $(0, \overline{q}]$. Both $\bar{\epsilon}$ and $\bar{q}$ are positive. To focus

[8] A difference to Sun's model is that in her model each buyer knows at which position he is located on the Hotelling line, but not where the product is located. My model can be thought of as a model where a product is always located in the middle of the Hotelling line, but buyers do not know where they are located before sampling or buying the product. Alternatively, the distance between a buyer and the product can be thought of as representing a match value, with a zero expectation ex-ante. This is similar to Bar-Isaac et al. (2010), where buyers receive with probability p a higher match value or with probability $1 - p$ a lower match value, where p is independent of buyers heterogeneous preferences for the quality of the product. Discrete choice demand estimation models (Berry et al., 1995) typically also capture consumer heterogeneity by assuming buyers have a match value for any given product that is distributed with a distribution that has a mean of zero.

on interior equilibria, I further assume that $\bar{q} = 3\bar{\epsilon}$. Further, I assume that the good exhibits characteristics of an experience good (Nelson, 1970). This assumption implies that a buyer can only learn his match ϵ_i after buying or sampling the product but not by consulting other sources of information. Quality q, on the other hand, can be known to a buyer before buying or sampling the product, depending on how much information is available to him from other sources. Marginal production costs are zero, an assumption that fits my empirical setting of a digital market.

I distinguish between two cases: Known and unknown quality. In the case of known quality, buyers know the product's quality q. This can be from reviews available from third parties, discussions in the press, word-of-mouth, online ratings, etc. The seller cannot influence how much of this type of information is available to buyers. She can, however, choose whether to disclose buyers' match value ϵ_i by offering a sample of her product. This captures the idea that it is easier to learn from information provided by third-parties about a product's vertical characteristics than about its horizontal characteristics.

In the case of unknown quality, buyers neither know the quality q of the product nor their match ϵ_i. The reason might be that the seller is an entrant without an established reputation or the product is new to the market. In this case, a seller can decide to disclose both the products quality q and buyers' match value ϵ_i by offering a sample.

Known Quality

In the case of known quality, the seller chooses a price for her product and whether to offer a sample to disclose buyers' match value ϵ_i. This is captured by a seller choosing a sampling strategy $S \in \{n, s\}$, where n stands for "not offering sample" and s for "offering sample." Based on her sampling strategy S, she chooses a corresponding price p_s^*.

If the seller does not offer a sample ($S = n$), buyers value the product uniformly at their expected utility which is equal to its known quality q. The profit maximizing price is

$$p_n^* = q.$$

The seller supplies all buyers, earning profits of

$$\Pi_n^* = q.$$

If she offers a sample ($S = s$), each buyer inspects the sample and learns his match value ϵ_i before buying the product. Buyers' valuations for the product are now dispersed and distributed uniformly in the interval $[q - \overline{\epsilon}, q + \overline{\epsilon}]$ resulting in a "rotated" (in the terminology of Johnson and Myatt, 2006) demand curve

$$D_s(p_s) = \frac{q - p_s + \overline{\epsilon}}{2\overline{\epsilon}} \quad (\text{for} \quad p_s \in [q - \overline{\epsilon}, q + \overline{\epsilon}]). \tag{1}$$

Profits are given by

$$\Pi_s = \frac{p_s(q - p_s + \overline{\epsilon})}{2\overline{\epsilon}} \quad (\text{for} \quad p_s \in [q - \overline{\epsilon}, q + \overline{\epsilon}]), \tag{2}$$

with the profit maximizing price given by

$$p_s^* = \frac{q + \overline{\epsilon}}{2}. \tag{3}$$

Equilibrium demand is

$$D_s(p_s^*) = \frac{\bar{\epsilon} + q}{4\bar{\epsilon}} \quad (\text{for} \quad p_s \in [q - \overline{\epsilon}, q + \overline{\epsilon}]) \tag{4}$$

Requiring that the quality of the highest type does not exceed $3\overline{\epsilon}$ ensures that (4) does not exceed 1. Notice also that a seller not offering a sample can supply all buyers at a price $p_n = q$, while a seller offering a sample will only supply half of the buyers at a price $p_s = q$. Therefore, a seller will only find it optimal to offer a sample if she can profitably raise her price, namely $p_s^* > p_n^*$. By doing so, she targets buyers who have a positive match value. Johnson and Myatt (2006) characterize this decision as the seller choosing either a "mass-market" position by keeping dispersion of buyers' valuations low, or choosing a "niche" position by increasing the dispersion of buyers' valuations to target buyers with a high valuation. A necessary condition for offering a sample, given any quality q', is $p_s^*(q') \geq p_n^*(q')$, which is necessarily implied by $D_s(p_s^*(q')) \leq D_n(p_n^*(q')) = 1$. This is summarized in my first proposition:

Proposition 1. *For any given quality known to buyers, offering a sample is associated with a higher (quality-adjusted) price.*

Firms equilibrium profits when offering a sample are

$$\Pi_s^* = \frac{(\bar{\epsilon} + q)^2}{8\bar{\epsilon}}. \tag{5}$$

The seller will choose to offer a sample if $\Pi_s^* > \Pi_n^*$, which is the case for

$$q < (3 - 2\sqrt{2})\bar{\epsilon}. \tag{6}$$

This implies that only sellers of a low quality offer a sample, a result already found by Sun (2011). The intuition is that sellers with a low quality have to ensure at least some buyers a good match. Therefore, they reveal their match to buyers' tastes by offering a sample. Sellers of a high quality, on the other hand, prefer to monetize their high quality, which is valued by all buyers. This can be done best by covering the whole market, i.e. to sell to as many buyers as possible. Offering buyers a sample and letting them learn their heterogeneous match value ϵ_i would be detrimental to this goal, as this would make buyers' valuations more disperse.

From this discussion, I take the following proposition:

Proposition 2. *When the quality of a product is known to buyers, a seller of a higher quality is less likely to offer a sample.*

Unknown Quality

In the case of unknown quality, the seller chooses whether to offer a sample to inform buyers about both her quality q and buyers' match value ϵ_i.

If a seller offers a sample ($S = s$), buyers learn her product's quality q and their match value ϵ_i for its horizontal characteristics. Demand D_s is given by the same expression as in the case of known quality by equation (1), the profit maximizing price p_s^* is given by equation (3), and equilibrium profits Π_s^* are given by equation (5).

Previously, a seller not offering a sample would earn a price equal to her true quality q. Now all sellers not offering a sample are pooled in a Bayesian equilibrium and earn a price equal to the average quality of all sellers not offering a sample in equilibrium. The price of sellers not offering a sample ($S = n$) is given by

$$p_n^* = E(q|S = n).$$

A seller choosing not to offer a sample supplies all buyers, implying

$$D_n(p_n^*) = 1,$$

earning profits of

$$\Pi_n^* = E(q|S = n). \tag{7}$$

This game has two types of Bayesian Equilibria: One where all sellers offer a sample (full unraveling) and one where only sellers of a high quality offer samples (partial unraveling). In the following, I concentrate on the more interesting case of the partial unraveling equilibrium.[9]

First note that the seller with the highest quality has always an incentive to offer a sample, implying that all sellers not offering a sample (no unraveling) is not an equi-

[9] In the full unraveling equilibrium, buyers' beliefs about the quality of sellers who do not offer a sample are sufficiently pessimistic such that in equilibrium all sellers offer a sample. This is the case if $E(q|S = n) < \frac{1}{8}\bar{\epsilon}$.

librium: If all sellers do not offer a sample, buyers' equilibrium expectation of a given seller's quality has to be equal to the average quality of all sellers in the market, namely $E(q|S = n) = \frac{1}{2}\bar{q} = \frac{3}{2}\bar{\epsilon}$. All quality types would earn profit of $\Pi_n = \frac{1}{2}\bar{q} = \frac{3}{2}\bar{\epsilon}$. But by plugging the quality of the highest type $\bar{q} = 3\bar{\epsilon}$ into equation (5), we see that the highest type would earn profits of $\Pi_s = 2\bar{\epsilon}$ when offering a sample. Therefore, she has an incentive to offer a sample. This implies that a potential partial unraveling equilibrium is where all types with a quality above a threshold q^* offer a sample, and all types with quality below q^* do not offer a sample. Buyers' equilibrium expectations of the quality of sellers not offering a sample is given by

$$E(q|S = n) = \frac{1}{2}q^*.$$

All sellers with $q < q^*$ are pooled at a price $p_n^* = \frac{1}{2}q^*$ and earn profits $\Pi_n^* = \frac{1}{2}q^*$.

To find q^*, we have to identify the type off seller who is indifferent between offering and not offering a sample, which is found by solving

$$\Pi_s^*(q = q^*) = \Pi_n^*(q = q^*) \tag{8}$$

for q^*. Inserting equations (2) and (7) into (8), and setting $q = q^*$ gives the unique solution of

$$q^* = \bar{\epsilon}. \tag{9}$$

In can be verified that for $q < \bar{\epsilon}$ it indeed holds that $\Pi_n^* > \Pi_s^*$ while for $q > \bar{\epsilon}$ it holds that $\Pi_n^* < \Pi_s^*$.

From this, I take the following proposition that I want to test empirically:

Proposition 3. *When the quality of a product is unknown to buyers, a seller of a higher quality is more likely to offer a sample.*

Looking at sellers' prices, we see that the seller of quality q^* sets a price of $p_s^* = \epsilon$, while all sellers not offering a sample set a price of $p_n^* = \frac{1}{2}\bar{\epsilon}$. As p_s^* is increasing in sellers' quality, all sellers offering a sample set higher prices than any seller not offering a sample. Therefore, even if quality is unobservable, we would observe that all sellers offering a sample set higher prices than sellers not offering a sample. This implies the following proposition:

Proposition 4. *When the quality of a product is unknown to buyers, offering a sample is associated with a higher price.*

Additionally, the fact that all sellers not offering a sample are pooled and set the same price $p_n^* = \frac{1}{2}\bar{\epsilon}$, irrespective of their quality, implies the following proposition:

Proposition 5. *Price dispersion is lower for sellers who do not offer a sample.*

Comparing the fraction of sellers who choose to offer a sample between the case of known quality, given by equation (6), with the case of unknown quality, given by equation (9), shows that less sellers offer a sample in the case of known quality. This proposition can be summarized as:

Proposition 6. *When the quality of products is unknown to buyers, there will be a larger fraction of sellers offering a sample than in the case of known quality.*

3 Empirical Setting

To examine the link between voluntary disclosure and pricing laid out in the theoretical propositions in the previous section, I use data collected from the independent e-book publishing platform *Smashwords.com*.

Self-Publishing and the Independent E-Book Market

Self-publishing by independent authors has in recent years evolved from a niche market to a market of substantial size. The growth has been driven by the substantial decrease in production and distribution cost brought by the internet, and the advent of e-readers and other digital devices suitable for reading.[10] Already in 2008, the yearly

[10]Similar developments in the movie and music industry have also been linked to a larger supply of new products (Waldfogel, 2015, 2012)

number of 275,000 self-published titles surpassed the number of books published by traditional publishers (Zindler, 2009). Since then, the number of self-published titles has steadily grown and reached 459,000 in 2013 (Bowker, 2014b). By 2013, self-published e-books accounted for an estimated tenth of both the number of books in bestseller lists and overall unit sales (Waldfogel and Reimers, 2015). While the median independent author still only earns between $500 - $1000 a year compared to $3,000 - $5,000 of a traditionally published author (Flood, 2015), in recent years there has been an increased professionalization among independent authors, with an industry expert commenting that there is a "trend of self-publisher as business-owner, rather than writer only" (Bowker, 2014a). There also have been some notable success stories among self-published authors, the most famous example being E.L. James's *50 Shades of Grey*, selling to-date more than 125 million copies worldwide (Stedman, 2015).[11]

Smashwords.com is the largest online distributor of self-published e-books. By 2012, it had become the top self-publisher on the internet with a total of yearly 90,252 published e-books (Bowker, 2013). E-books that are published on *Smashwords* are not only sold on its website but also distributed to the largest online e-book sellers such as *Apple's iBookstore*, *Barnes & Noble*, or *Kobo*, although not to *Amazon*, which has its own self-publishing program. While most *Smashwords* titles are written by amateur authors, who at best serve only small niche markets, some titles have also been very successful and even reached international bestseller status, for example on *Apple's iBookstore* (Coker, 2012). Because *Smashwords* supplies a large share of the titles available on the *iBookstore*, *Forbes.com* termed it "Apple's biggest (unknown) supplier of e-books" (Colao, 2012).

On *Smashwords*, authors have full control over most aspects of how their e-books are sold. Most important for my study, authors themselves set the retail price at which their e-books are sold on the *Smashwords* website and on the other platforms their e-books are distributed to. Of the retail price, *Smashwords* earns a commission between

[11]Interestingly, the author of *50 Shades of Grey* signed with a major publisher once the series had become an underground self-published success. Such a strategy seems not to be uncommon among successful independent authors.

10 and 20 percent. Depending on whether the e-book is sold directly on *Smashwords* or on a platform it is distributed to, which each take their sales commission, authors earn between 50 and 80 percent of the retail price. Additional to its price, authors can choose whether to offer a free sample of their e-book, which can consist of any length (0-100 percent) of the full e-book.

The following features make the market for e-books and *Smashwords*, in particular, interesting for the study of disclosure and pricing: E-books can be considered a typical case of experience goods (Nelson, 1970) since it is difficult for readers to know their utility before reading an e-book. Therefore, offering a free sample of an e-book should be a particularly convenient option for an author to reveal the characteristics of her e-book. Offering a sample is not only costless since the sample is provided digitally, but it also has a high degree of credibility, as readers can sample part of the "real" product, instead of relying on third-party information such as reviews. Another feature of this setting is that authors can choose for themselves whether to offer a sample, which for instance is not the case for an author published by a traditional publisher or on *Amazon's* self-publishing platform, where offering a sample is mandatory.[12] While *Smashwords* does not guarantee that when distributing an e-book to another platform the sample size will be the same as on its own platform, it states that it will make every effort to communicate to authors differing sampling practices and give authors the opportunity to opt out of any outlet in which the sampling practice does not meet the authors approval.[13]

Another feature is that the model in the previous section assumes that the seller is a monopolist. Although I cannot fully rule out strategic interactions between authors in terms of pricing and offering a sample, the fact that there are more than 300,000 e-books offered on *Smashwords* makes it unlikely that there is a direct strategic interaction between particular authors. This and the fact that prices are positive despite zero marginal costs suggest that the e-book market resembles a case of monopolistic

[12]When publishing an e-book on *Amazon* a 10 percent sample of each e-book is mandatory. Source: https://kdp.amazon.com/help?topicId=A25WS075EUM6NF (Accessed January 12, 2014).

[13]Source: https://www.smashwords.com/about/supportfaq (Accessed November 12, 2015).

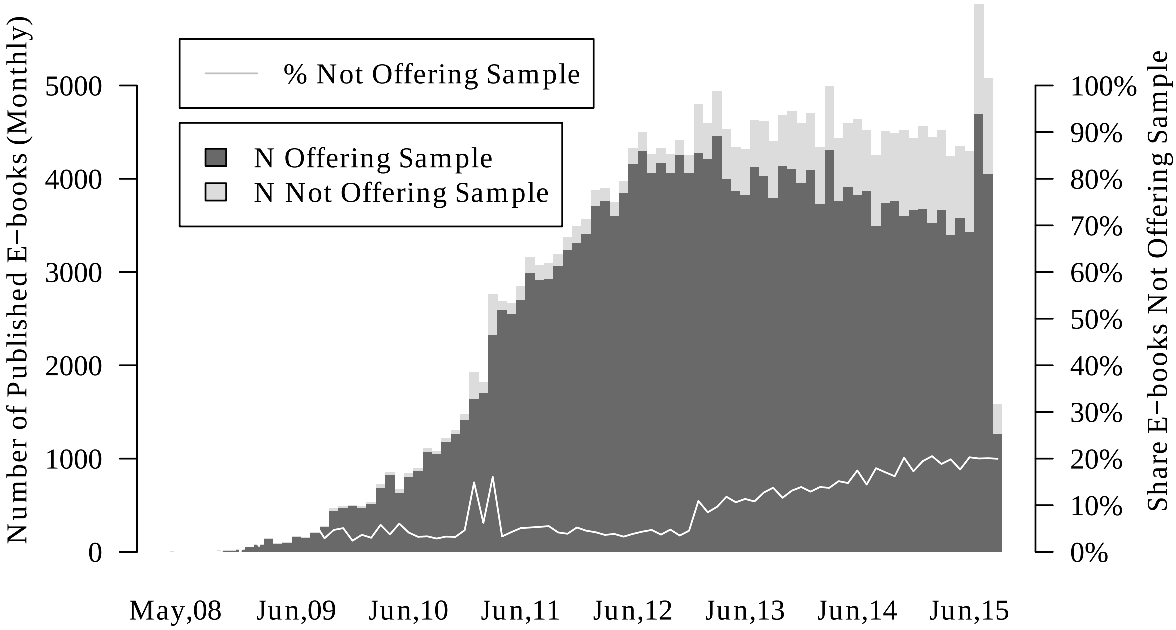

Figure 1: *Monthly number of e-books published on Smashwords.com and share of e-books not offering a sample.*

competition (Chamberlin, 1933), with each author optimizing in view of a demand curve determined by the collective decisions of all other authors offering e-books on *Smashwords*. It also has to be mentioned that since the entry cost into self-publishing is virtually zero and *Smashwords* does not have any kind of quality control for e-books published on their website, it is very likely that a large number of low-quality authors enter the market who fail to meet basic standards of writing such as correct spelling or grammar.

Dataset

I have collected via web-scraping an almost complete sample of e-books offered on *Smashwords.com*. I collected the data between September 12 and September 22, 2015, obtaining data on 367,000 of the total 375,000 e-books ever published on *Smashwords*. In the following analysis, I exclude 55,000 e-books that are offered for free and 200 e-books that are sold at a price above $50. I also exclude 65,000 e-books of authors that have more than 50 e-books on offer, who represent only 0.4% of authors publishing on

Category	Number Books	%	Subcategory	Number Books	%
Fiction	182,028	72.7	Erotica	38,357	13.2
Non-Fiction	61,251	26.3	Romance	29,614	10.2
Essay	1,892	0.7	Fantasy	15,087	5.2
Screenplays	540	0.2	Science Fiction	11,860	4.1
Plays	448	0.2	Young Adult or Teen	9,569	3.3
			Mystery and Detective	9,203	3.2
			Thriller and Suspense	8,123	2.8
			Religion and Spirituality	6,962	2.4
			Literature	7,180	2.5
			(Others)	110,204	41.6
Total	246,159	100.0	Total	246,159	100.0

Table 1: *Summary of published genres on Smashwords.*

Smashwords but supply 16 percent of all e-books.[14] I exclude these e-books as they have a disproportional impact on my summary statistics and graphs. In section 5, however, I show that they do not have a substantial impact on my regression results. In the end, I arrive at a sample size of 246,157 e-books I use in my analysis.

Figure (1) shows the monthly number of published e-books starting with April 2008, when the platform first went online. The website experienced steady growth in terms of published e-books in the years 2008 to 2012, and stabilized in 2012 at between 5,000 and 6,000 monthly published e-books. While for e-books published in the earlier years, 2008 - 2012, almost all authors offer a sample, for e-books published in 2013 and later there is a small trend towards not offering a sample. Of the e-books published in 2015, approximately 80 percent offer a sample. According to table (1), most e-books on *Smashwords* are fiction. Erotica and Romance are the most popular genres, which is in accordance with the general trend in e-publishing.[15]

Table (2) and (3) describe and summarize the variables contained in my dataset. The two variables that are most important for my analysis are the price of each e-book (in U.S. Dollars) and the sample size. Authors can set any price equal or above $0.99 for their e-book, and choose any sample size between 0 and 100 percent of an

[14]The most extreme case is one author publishing 17,000 musical scores of well-known tunes and classical pieces.

[15]The commonly cited explanation for the over-representation of such genres as Romance and Erotica is that e-readers offer a better anonymity by better hiding what readers are reading. As a (female) literature critic put it in *The Guardian*: "The reading public in private is lazy and smutty. E-readers hide the material. . . . My own . . . literary fetish is male-oriented historical fiction . . . I'm happier reading it on an e-reader, and keeping shelf space for books that proclaim my cleverness." (Senior, 2012)

e-book. The distribution of prices is skewed, with a mean price of \$3.73, a median price of \$2.99, and a maximum price of \$50. Therefore, in my analysis, I use the natural logarithm of price. For the average e-book 17 percent of its full length are offered as a sample, while overall for 89 percent of e-books a sample is offered. As the sample size is to a large degree determined by the length of an e-book (correlation -0.08 between sample size as a percentage of the whole e-book and length of the e-book in words), I will in most parts of my analysis use a dummy indicating whether a sample is offered for a given e-book.[16] For a sub-group of 25,880 e-books, ratings are directly posted by previous buyers[17] on *Smashwords*, while for 15,817 e-books I additionally collected ratings from the social reading community *Goodreads.com*,[18] and for 31,106 e-books I collected ratings posted on *Amazon.com*. For the average e-book, there are about 9 other e-books published by the same author on *Smashwords*, while the median e-book has 4 other e-books published by the same author. In my analysis, I therefore use the natural logarithm of this variable, as I do for all other similarly skewed variables.

4 Results

I organize my results into three parts: In section 4.1, I examine the empirical relationship between offering a sample and an e-book's price, therefore testing propositions 1 and 4. In section 4.2, I show which characteristics of an e-book make an author more likely to offer a sample, therefore testing propositions 2, 3 and 6. In section 4.3, I examine the relationship between offering a sample and price dispersion, thus testing proposition 5.

[16]Using the continuous variable for sample size produces qualitatively the same results but with coefficients that are less straightforward to interpret.

[17]Only verified buyers can rate an e-book on *Smashwords*.

[18]*Goodreads.com* is the largest online reading community, where readers can post online reviews, connect with other readers and browse the catalog of books to discover new books. Goodreads boast as having 40 million active members, a catalog of 1.1 billion books with 43 million reviews (Source: http://www.goodreads.com/about/us; accessed October 25, 2015). In 2013, *Goodreads* was acquired by *Amazon*.

Variable	Description
Price	E-book's price in U.S. Dollars
Sample Size (Percent)	Size of sample in percent of e-book
Sample Offered (Yes/No)	Dummy indicating whether a sample is offered
Rating Available (Yes/No) (Smashw.)	Dummy indicating whether e-book is rated on *Smashwords.com*
Rating Available (Yes/No) (Goodreads)	Dummy indicating whether e-book is rated on *Goodreads.com*
Rating Available (Yes/No) (Amazon)	Dummy indicating whether e-book is rated on *Amazon.com*
N Ratings (Smashw.)	Number of ratings an e-book has received on *Smashwords.com*
N Ratings (Goodreads)	Number of ratings an e-book has received on *Goodreads.com*
N Ratings (Amazon)	Number of ratings an e-book has received on *Amazon.com*
Average Rating (Smashw.)	Average rating of an e-book on *Smashwords.com*
Average Rating (Goodreads)	Average rating of an e-book on *Goodreads.com*
Average Rating (Amazon)	Average rating of an e-book on *Amazon.com*
Category	Categorical variable indicating category (genre) of e-book
Subcategory	Categorical variable indicating subcategory (sub-genre) of e-book
Language	Categorical variable indicating language of e-book
Gender Author	Probability that author is female as implied by first name
Year Published	Categorical variable indicating year e-books has been published on *Smashwords.com*
Length E-book in Words	Total length of e-book in words (divided by 1000)
Time Since Published (Weeks)	Number of weeks since an e-book has been published on *Smashwords.com*
Number Previous E-books	Number of e-books same author has previously published on *Smashwords.com*
Previous E-books Available (Yes/No)	Dummy indicating whether previous e-books by same author are available on *Smashwords.com*
Number E-books	Total number of e-books same author has published on *Smashwords.com*
Previous Ratings Available (Yes/No)	Whether ratings are available on *Smashwords.com* for previous e-books by same author
Number Ratings Previous E-books	Number of ratings that are available on *Smashwords.com* for previous e-books of same author
Average Rating Previous E-Books	Average rating on *Smashwords.com* for previous e-books of same author

Table 2: *Description - Main Variables and Controls*

Statistic	N	Mean	St. Dev.	Min	Median	Max
Price	246,157	3.73	2.99	0.99	2.99	50.00
Sample Size (Percent)	246,157	16.68	9.83	0	20	100
Sample Offered (Yes/No)	246,157	0.89	0.31	0	1	1
Rating Available (Yes/No) (Smashw.)	246,157	0.11	0.31	0	0	1
Rating Available (Yes/No) (Goodreads)	246,157	0.06	0.25	0	0	1
Rating Available (Yes/No) (Amazon)	246,157	0.13	0.33	0	0	1
N Ratings (Smashw.)	246,157	0.27	1.58	0	0	209
N Ratings (Goodreads)	246,157	58.73	4,081.00	0	0	1,770,332
N Ratings (Amazon)	246,157	2.42	20.49	0	0	997
Average Rating (Smashw.)	25,880	4.53	0.73	1.00	5.00	5.00
Average Rating (Goodreads)	15,817	3.87	0.52	1.00	3.90	5.00
Average Rating (Amazon)	31,106	4.37	0.72	1.00	4.50	5.00
Length E-Book in Words (Words/1000)	246,157	43.86	48.36	0.01	28.30	1,990.00
Time Since Published (Weeks)	246,157	122.40	75.66	1.14	117.70	384.40
Previous E-books Available (Yes/No)	246,157	0.65	0.48	0	1	1
Number Previous E-books	246,157	4.62	7.21	0	2	49
Number E-books	246,157	10.15	11.38	1	5	50
Previous Ratings Available (Yes/No)	246,157	0.26	0.44	0	0	1
Number Ratings Previous E-books	246,157	2.08	11.48	0	0	1,311
Average Rating Previous E-Books	61,573	4.33	0.80	1.00	4.50	5.00

Table 3: *Summary Statistics - Main Variables and Controls*

4.1 Relationship between Offering Sample and Price

According to proposition 1 and 4, offering a sample should be associated with a higher price. In the case of known quality (proposition 1), the positive association should be driven by a sample increasing the match between buyers' tastes and e-books' horizontal characteristics, resulting in a match-premium in the form of a higher price. In the case of unknown quality (proposition 4), the positive association should be driven by authors of better quality e-books being more likely to offer a sample and earning a premium based on their better quality. Authors of lower quality e-books not offering a sample should be pooled at a low price reflecting their lower quality.

In my empirical analysis, I distinguish between known and unknown quality by differentiating between e-books with and without posted online ratings. As two alternative measures for how much is known about quality, I use whether authors release their first e-book and whether previous e-books of the same author have been rated. In section 5, I use ratings collected from two other online platforms, *Goodreads.com* and *Amazon.com*, as alternative measures for the availability of information on quality.

The availability of a posted online rating is a good measure of whether quality is known to prospective buyers for the following reason: Quality is defined by economists as any characteristic that is positively valued by every buyer. Higher quality therefore monotonically increases every buyers' utility. If an e-book possesses higher quality,[19] this should be reflected in a higher average rating.[20] Characteristics that are only valued by some buyers, for example, a particular humor, literary style, slang or dialect, can be thought of as horizontal characteristics. Such horizontal characteristics should increase the dispersion of ratings but not the average.[21] Therefore, if a buyer observes the average online rating of an e-book, this reveals to him information about an e-book's quality. In contrast to quality, it should be much harder for a buyer to learn about an e-book's horizontal characteristics by examining online ratings. To do this, a buyer would need to carefully examine each review, including the posted text, and judge which particular horizontal characteristic made a previous reader assign the e-book a high or a low rating.

In the two upper panels of figure (2), the relationship between the sample size as a fraction of an e-book and the logarithm of its price is shown. On the upper panel, the relationship is shown for e-books without a posted rating on *Smashwords*. On the middle panel, the relationship is shown for e-books with a posted rating. In both panels, given that a (positive size) sample is offered for an e-book, the average price declines as the sample size increases. In a separate regression, I found that this pattern is largely explained by shorter e-books, which are on average priced lower, usually offering a larger sample in terms of percentage of the whole e-book. Therefore, I will use in my analysis a dummy indicating whether any positive sample size is offered for a given e-

[19] As a researcher, I am agnostic towards what characteristics constitute "high quality." Online ratings can, therefore, be thought of as a "reader-centric" measure of "quality" as opposed to characteristics that are valued by more sophisticated readers such as literary critics. For example, in the data that I have collected from the book review website *Goodreads.com*, the list of best-rated books is dominated by authors of popular bestseller such as J.K. Rowling, George R.R. Martin, or Stephen King, and not the likes of Dostoyevsky, Thomas Mann, or Marcel Proust.

[20] Using a seller's rating as a measure of her quality is a common approach in many studies on internet markets such as e-Bay. An exception is Jin and Kato (2006), who purchase baseball cards on e-Bay and let a professional grade these cards. They find that sellers with better rating are indeed more likely to deliver and less likely to sell counterfeited cards, although conditional on delivery, they do not deliver a better quality.

[21] Sun (2012) employs the same argument and uses the dispersion of online ratings as a measure for horizontal location of a product.

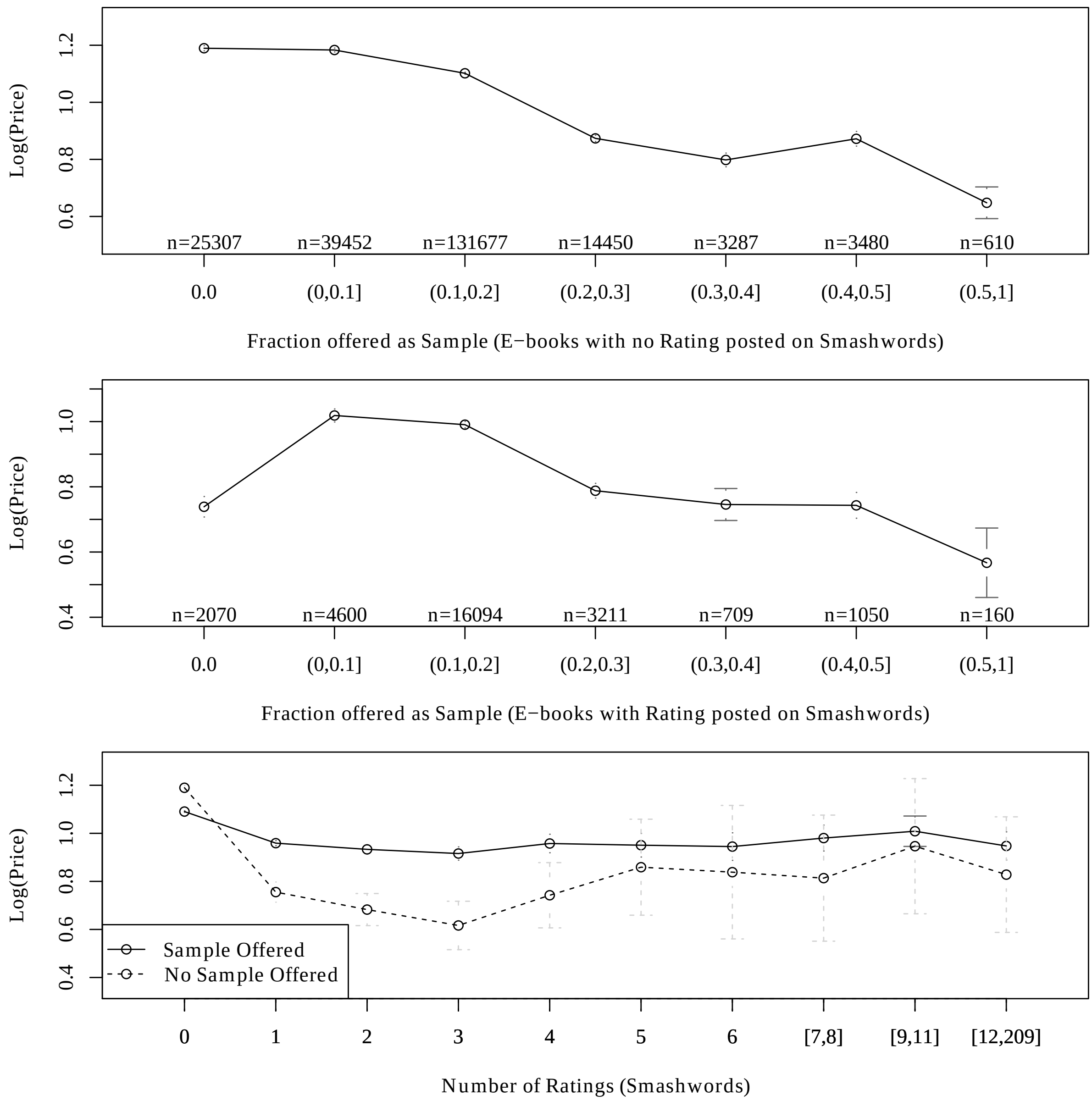

Figure 2: *On the upper two panels, the mean of log-price of e-books is plotted against the fraction offered as sample. On the lower panel, the mean of log-price of e-books is plotted against the number of ratings on* Smashwords*. Bars show 95 percent confidence intervals. If a confidence interval is not shown, it is too narrow to be visible on the plot.*

Sample: E-Books without rating	Mean Price	Mean Log(price)	N	Share
Sample Offered = No	4.30	1.19	25,420	11.5 %
Sample Offered = Yes	3.74	1.09	194,857	88.5 %
Difference	0.55	0.10		
p-value (H_0 : Equal Means)	<0.001	<0.001		
Sample: E-Books with rating	Mean Price	Mean Log(price)	N	Share
Sample Offered = No	2.76	0.72	1,957	7.6 %
Sample Offered = Yes	3.15	0.94	23,923	92.4 %
Difference	-0.40	-0.22		
p-value (H_0 : Equal Means)	<0.001	<0.001		

Table 4: *Means of e-books' price by sub-groups of e-books offering or not offering a sample, and e-books with or without posted online ratings. Statistical test for equal means is based on a Welch two-sample t-test assuming unequal variances.*

book. Using the continuous sample size leads to qualitatively similar results but makes the interpretation of the coefficients more complicated.

The lower panel of figure (2) shows the relationship between the average of the logarithm of price and the number of ratings an e-book has received. The sample is split into e-books that offer a sample (solid line) and e-books that do not offer a sample (dashed line). In the case where e-books are rated at least once, e-books for which a sample is offered have a higher average price than e-books that do not offer a sample. This empirical relationship is consistent with the positive relationship predicted by proposition 1. In the case where e-books have not been rated, the relationship is reversed: e-books for which a sample is offered have a lower average price than e-books for which a sample is not offered. This is the opposite of the positive relationship between disclosure and price predicted by proposition 4.

Comparing average prices and testing for the statistical significance of the observed differences in table (4), the same basic pattern can be observed: In the case of rated e-books, e-books for which a sample is offered have a statistically significantly lower price (p-value < 0.001). In the case of rated e-books, e-books for which a sample is offered have a statistically significant higher price (p-value < 0.001).

The same relationship can also be observed in a set of regressions shown in table

	Dependent variable:			
	Log(Price)			
	(1)	(2)	(3)	(4)
Sample Offered (Yes/No)	−0.063***	−0.078***	−0.117***	−0.110***
	(0.004)	(0.004)	(0.007)	(0.007)
Log(1 + N Ratings) (Smashw.)		−0.050***	−0.052***	−0.047***
		(0.007)	(0.007)	(0.007)
Rating Available (Yes/No) (Smashw.)		−0.290***	−0.306***	−0.244***
		(0.016)	(0.016)	(0.016)
Average Rating (Smashw.)		0.027***	0.029***	0.031***
		(0.005)	(0.005)	(0.005)
Previous Book Available (Yes/No)			−0.153***	−0.124***
			(0.009)	(0.010)
Rating Available Previous Book (Smashw.)				−0.176***
				(0.010)
Sample Offered X Rating Available (Smashw.)		0.185***	0.198***	0.164***
		(0.015)	(0.015)	(0.016)
Sample Offered X Previous Book Available			0.050***	0.036***
			(0.009)	(0.010)
Sample Offered X Rating Available Previous Book				0.065***
				(0.011)
Constant	0.753***	0.721***	0.769***	0.740***
	(0.072)	(0.070)	(0.067)	(0.070)
Category (Categorical)	Yes	Yes	Yes	Yes
Subcategory (Categorical)	Yes	Yes	Yes	Yes
Language (Categorical)	Yes	Yes	Yes	Yes
Gender Author (Categorical)	Yes	Yes	Yes	Yes
Year Published (Categorical)	Yes	Yes	Yes	Yes
Length E-Book in Words	Yes	Yes	Yes	Yes
Time Since Published (Weeks)	Yes	Yes	Yes	Yes
Number Previous Books	Yes	Yes	Yes	Yes
Number Books	Yes	Yes	Yes	Yes
Observations	246,157	246,157	246,157	246,157
R^2	0.245	0.251	0.240	0.257
Note:				*p<0.1; **p<0.05; ***p<0.01
				Robust standard errors in parantheses

Table 5: *Cross-sectional linear regressions with an e-book's price as the dependent variable and an individual e-book as the unit of observation.*

(5), where I control for the observed characteristics of each e-book.[22] In column (1), I do not differentiate between rated and not-rated e-books. In this regression, offering a sample is associated with a 6.3 percent lower price (p-value < 0.01). In column (2), I differentiated between e-books with a known and unknown quality by including a dummy indicating whether a given e-book has been rated. I include this dummy as a main effect and as an interaction with the variable indicating whether a sample is offered. I also include the total number of ratings as a separate continuous variable. With the dummy, I capture the discontinuity between zero and one rating. The lower panel in figure (2) already visually hints at the presence of such a discontinuity in the relationship between the number of ratings and price. Additionally, I include the average rating for e-books that have been rated at least once.[23] In this regression, the main effect of offering a sample is negative, as in column (1). However, the interaction between whether an e-book is rated and whether a sample is offered is positive. The size of the coefficients indicates that non-rated e-books that offer a sample have a 7.8 percent lower price (p-value < 0.01). E-books that are rated and offer a sample, on the other hand, have an 18.5 percent higher price (p-value < 0.01). The quality of an e-book as measured by its average rating has also a significant impact on price, although the size of this effect is not large: One additional "star" in rating is associated with a 2.7 percent higher price (p-value < 0.01).[24]

Interestingly, the main effect of being rated indicates that being rated is associated with a 29.0 percent lower price (p-value < 0.01) in the regression in column (2). There are two possible explanations for this result: One is that the effect is driven by a re-

[22]At this point, I have to note that in these regressions I am only testing for a correlation between offering a sample and an e-book's price while controlling for other potentially confounding factors that I observe in my data. This, however, is in line with both the theoretical model, where price and offering a sample is a simultaneous choice, and my empirical setting, where authors can change both variables at any point in time within the same online form. The other observable characteristics of an e-book, on the other hand, can be considered to be exogenous in the sense that they are already determined at the time when the author chooses the price and the sample size of her e-book.

[23]The variable *Average Rating* is demeaned in column (2) - (4) and filled up with zeros in case no rating is available. Therefore, it is mathematically equivalent to an interaction between the average rating and the dummy indicating whether a rating is available. This method of dealing with missing values is discussed e.g. in Gelman and Hill (2006).

[24]Cabral and Hortacsu (2010) similarly find in a study on eBay sellers that in a cross-sectional regression there is only a weak relationship between a seller's reputation measure and her sales price.

verse causality, as e-books with a higher price may be less likely to be rated due to their lower sales. The other explanation is that authors of e-books with an unobserved quality inflate their prices. This explanation would be consistent with similar evidence from the lab showing that sellers inflate their prices in treatments with unobservable product quality (Henze et al., 2015). To rule out that the first explanation is the main driver behind this result, in the next section I use an instrumental variables approach, instrumenting for the availability of a rating by using how long an e-book has been on sale on *Smashwords*.

In column (3) and (4), I use two alternative measures of whether information about an author's quality is known. In column (3), I add a dummy indicating whether an author has released any previous e-books. In column (4), I add a dummy indicating whether previous e-books of an author have been rated. I interact both dummy variables with the dummy indicating whether a sample is offered. I also add the number of e-books previously released by an author and the number of ratings previous e-books of the same author have received as continuous variables. Using both alternative measures of whether the quality of an author is known produces similar results as using the availability of a rating.

Instrumenting for Rating Availability

A potential concern with the previous regression analysis is that the regression equation includes price as the dependent variable while the independent variable indicating whether an e-book has been rated should be positively correlated with unobserved demand for an e-book. This may introduce an omitted variable bias into my previous estimates.[25]

To control for this potential source of bias, I use the number of weeks an e-book has been on the market as an instrument for whether it has been rated. The following arguments indicate that this is a valid instrument: An e-book is more likely to

[25]However, as the correlation between price and unobserved demand is picked up by the coefficient on the dummy indicating whether a rating is available, the coefficient on the interaction between rating availability and offering a sample should not be significantly impacted by this correlation. See Bun and Harrison (2014) for a discussion of regression models that include endogenous interaction effects.

be rated the longer it has been on sale on *Smashwords*. Indeed, there is a strong correlation of $+0.18$ between the availability of ratings and the number of weeks since an e-book has been published. An argument for the exogeneity of this instrument is that whereas for physical books retailers have an incentive to lower prices to sell-off older books to free shelf-space, this is not an issue in the case of e-books. For example, whereas Aguzzoni et al. (Forthcoming) find a negative relationship between the time a book has been on the market and its price in the case of physical books; De los Santos, Babur and Wildenbeest (2015) find no statistically significant relationship in the case of e-books sold on *Amazon.com*. An additional argument is that the sales of self-published e-books are more driven by word-of-mouth effects than traditional books, as self-published books usually do not profit from advertising campaigns as many traditionally published books do. Self-published e-books should, therefore, have more evenly spread-out sales over time, as is usual for books whose sales are driven by Word-of-Mouth (Beck, 2007). This reduces incentives to employ intertemporal price discrimination strategies.[26]

Table (6) shows the results of my instrumental variables regressions. Additionally to instrumenting for the availability of ratings, I instrument for the interaction between the availability of ratings and whether a sample is offered, the average rating, and the number of ratings by including interactions between these variables and the number of weeks that have passed as instruments. All instruments are statistically significant in the reduced form regressions in column (1) and (2) in table (6). They are also statistically significant in the first stage regressions, which can be found in table (14) in the appendix. Column (3) and (4) in table (6) show the results of two instrumental variable regressions. In column (3), only the endogenous variable indicating whether a rating is available is included in the regression. In column (4), I add the number of ratings of a given e-book as an endogenous variable. This regression is closed to the specifica-

[26] I also explored the use of the aggregated number of Google Searches for the term "Smashwords" as an instrument. This should be a measure for the internet traffic *Smashwords* receives and, therefore, increase the likelihood that an e-book has been rated. Unfortunately, this variable is almost fully explained by the number of weeks that have passed since an e-book has been published, indicating that there is not sufficient variation in the number of aggregated searches to use it as an instrument.

	Dependent variable: Log(Price)			
	OLS Reduced Form		instrumental variable	
	(1)	(2)	(3)	(4)
Sample Offered (Yes/No)	-0.0623^{***}	-0.1344^{***}	-0.1345^{***}	-0.1446^{***}
	(0.0040)	(0.0065)	(0.0067)	(0.0081)
Rating Available (Yes/No) (Smashw.)			-0.7335^{***}	1.2870^{***}
			(0.0689)	(0.3085)
Sample Offered X Rating Available			0.9384^{***}	0.9966^{***}
			(0.0703)	(0.0824)
Average Rating (Smashw.)			0.0302^{***}	0.1024^{***}
			(0.0057)	(0.0126)
Log(1 + N Ratings) (Smashw.)				-1.6100^{***}
				(0.2173)
Time Since Published (Weeks)	0.00005^{***}	-0.0005^{***}		
	(0.00002)	(0.0001)		
Time Since Published X Sample Offered		0.0008^{***}		
		(0.0001)		
Time Since Published X Average Rating		0.0002^{***}		
		(0.00003)		
Time Since Published X Log(1 + N Ratings)		-0.0005^{***}		
		(0.00002)		
Constant	0.6817^{***}	0.7316^{***}	0.7486^{***}	0.7589^{***}
	(0.0180)	(0.0186)	(0.0189)	(0.0219)
Category (Categorical)	Yes	Yes	Yes	Yes
Subcategory (Categorical)	Yes	Yes	Yes	Yes
Language (Categorical)	Yes	Yes	Yes	Yes
Gender Author	Yes	Yes	Yes	Yes
Length E-Book in Words	Yes	Yes	Yes	Yes
Number Previous Books	Yes	Yes	Yes	Yes
Number Ratings Previous Books	Yes	Yes	Yes	Yes
Number Books	Yes	Yes	Yes	Yes
Stock-Yogo Weak Identification Test (F-Stat.)			1444.2	54.1
Observations	246,157	246,157	246,157	246,157
R^2	0.2482	0.2510	0.2317	-0.0134

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$
Robust standard errors in parantheses

Table 6: *Instrumental variables regressions using the number of weeks an e-book is on the market as an instrument for whether a rating is available. The dependent variable is the logarithm of price of a given e-book.*

tion in the previous section in table 5. However, the F-test for joint weak identification indicates that there is not sufficient variation in my instruments to jointly identify the causal effect of all four instrumented-for variables (F-value=54.1). Therefore, the regression in column (3) is my preferred specification.

The results of the instrumental variables regressions are similar to the results of my previous regressions in table (5). The coefficients for the availability of a rating, for whether a sample is offered, and for the interaction between rating availability and whether a sample is offered are statistically significant and have the same signs as in table (5). For e-books that are not rated, offering a sample is associated with a lower price, while for e-books that are rated, offering a sample is associated with a higher price. The main effect of being rated on price is also negative in the regression in column (3). In column (4) it is positive, but the test for joint weak identification indicates that there is not sufficient variation to identify the discontinuity between zero and one ratings by including both the dummy the continuous measure of the number of ratings an e-book has received.

Overall, I find strong support for proposition 1, suggesting that authors of e-books where more is known about their quality earn a higher price when disclosing their horizontal characteristics by offering a sample. I interpret this in favor of the hypothesis that disclosing information on horizontal product characteristics increases the match between products and buyers' tastes, enabling the seller to earn a higher price. In contrast, I find support against proposition 4, suggesting that sellers disclosing product information when less is known about their quality do not earn a higher price compared to non-disclosing sellers. In fact, sellers who disclose by offering a sample earn a lower price. In section 6, I show how my results can be explained by incorporating a fraction of naive buyers into the model.

4.2 For Which Type of E-book Samples are Offered

Propositions 2, 3, and 6 give three predictions on the general level of disclosure and what type of author discloses by offering a sample: According to proposition 6, more authors should disclose their characteristics by offering a sample when the quality of their e-books is unknown, i.e. when their e-books are not rated. According to proposition 3, when quality is unknown, i.e. e-books are not rated, authors of high-quality e-books should be more likely to disclose their characteristics by offering a sample than authors of a low-quality e-book. When quality is known i.e. for rated e-books, according to proposition 2 this association should be reversed, as authors of high-quality e-books should be less likely to disclose by offering samples.

As in the previous part, I distinguish between known and unknown quality by looking at whether an e-book has been rated on *Smashwords*. In section 5, I present results using rating availability on *Goodreads* and *Amazon* as a robustness check, obtaining similar results. Additionally to the availability of rating, I also use the number of previously published e-books and the number of ratings for previously published e-books as measures of whether quality is known. As a measure of quality, I use ratings for a given e-book and ratings for previous e-books of the same author published on *Smashwords*.

Figure (3) shows three plots of the relationship between the fraction of e-books offering a sample and the three different measures of availability of information on quality. In the upper panel, the relationship with the number of ratings a given e-book has received is shown. Visually, there is a discontinuity between e-books that have no rating and e-books that have one rating: E-books with one rating are about 4 percent more likely to offer a sample. For e-books that have more than one rating, the probability of offering a sample stays approximately constant. For the number of previously published e-books and the number of ratings of previously published e-books, the pattern looks similar, although it is less pronounced. Overall, this is first evidence against proposition 6.

In table (2), I show results from a series of logit regressions with the probability that an e-book offers a sample as the dependent variable. In general, the results confirm the pattern observed in figure (3): In column (2), the dummy indicating whether a given

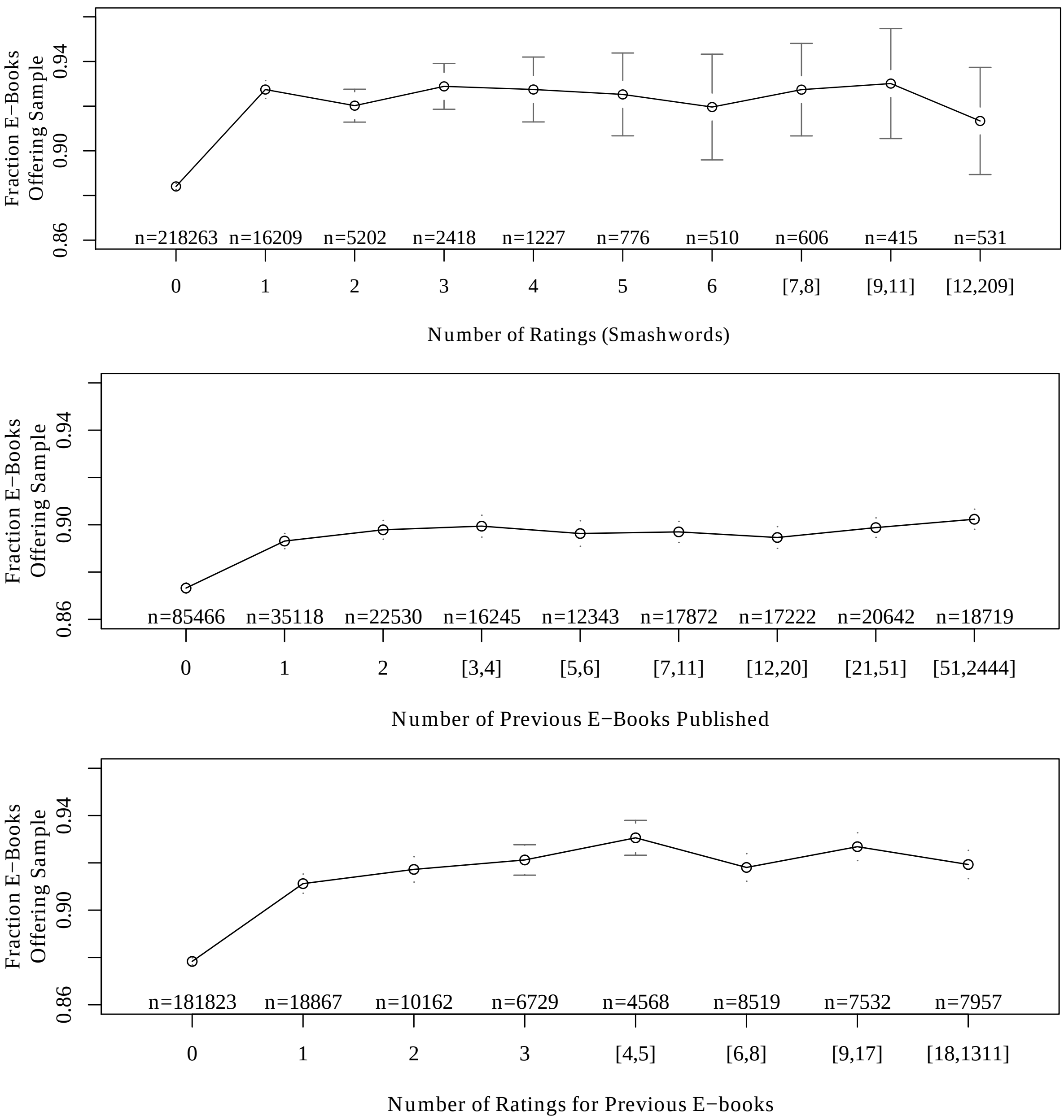

Figure 3: *Three different measures of how much information is known about a given e-book plotted against the fraction of e-books offering a sample. Bars show 95 percent confidence intervals. If a confidence interval is not shown, it is too narrow to be visible on the plot.*

	Dependent variable:			
	Sample Offered (Yes=1)			
	(1)	(2)	(3)	(4)
Log(Price)	-0.013^{***}	-0.013^{***}	-0.013^{***}	-0.012^{***}
	(0.001)	(0.001)	(0.001)	(0.001)
Rating Available (Yes/No) (Smashw.)		0.009^{**}	0.009^{**}	0.007^{*}
		(0.004)	(0.004)	(0.004)
Log(1 + N Ratings) (Smashw.)		-0.013^{***}	-0.012^{***}	-0.012^{***}
		(0.004)	(0.004)	(0.004)
Average Rating (Smashw.)		0.022^{***}	0.022^{***}	0.021^{***}
		(0.002)	(0.002)	(0.002)
Previous E-books Available (Yes/No)			0.004^{**}	0.003
			(0.002)	(0.002)
Previous Ratings Available (Yes/No)				0.010^{***}
				(0.002)
Average Rating Previous E-books				0.004^{**}
				(0.002)
Constant	0.213^{**}	0.215^{**}	0.209^{**}	0.210^{**}
	(0.083)	(0.083)	(0.083)	(0.082)
Category (Categorical)	Yes	Yes	Yes	Yes
Subcategory (Categorical)	Yes	Yes	Yes	Yes
Language (Categorical)	Yes	Yes	Yes	Yes
Gender Author (Categorical)	Yes	Yes	Yes	Yes
Year Published (Categorical)	Yes	Yes	Yes	Yes
Length E-Book in Words	Yes	Yes	Yes	Yes
Time Since Published (Weeks)	Yes	Yes	Yes	Yes
Number Previous Books	No	No	Yes	Yes
Observations	246,157	246,157	246,157	246,157

Note: *p<0.1; **p<0.05; ***p<0.01
Coefficients give marginal effects estimates; standard errors in paranthenses

Table 7: *Cross-sectional logit regressions with the probability that an e-book offers a sample as the dependent variable and an individual e-book as the unit of observation.*

e-book has been rated on *Smashwords* is positive and statistically significant (p-value < 0.05). However, the effect is smaller than figure (3) suggests, with an e-book for which a rating is available being 0.9 percent more likely to offer a sample. In column (3), I add a dummy indicating whether previously published e-books by the same author are available on *Smashwords*. The coefficient on the dummy is also positive and statistically significant (p-value < 0.05), indicating that authors with previously published e-books are 0.4 percent more likely to offer a sample for a given e-book. In column (4), I add a dummy indicating whether a previous e-book of the same author has received a rating. If a previous e-book has been rated, the probability of offering a sample increases by 1 percent (p-value < 0.01). Overall, I find evidence against proposition 1, as authors of e-books where more is known about their quality are more likely to disclose by offering a sample.

To examine the prediction of proposition 2 that in the case of known quality authors of better quality e-books should be less likely to offer a sample, I look at whether e-books with a higher quality are less likely to offer a sample. As a measure of known quality, I use the average rating of an e-book posted on *Smashwords.* In column (2), (3) and (4), the coefficient on the average rating is positive and statistically significant (p-value < 0.01), indicating that an additional "star" in rating increases the probability that a sample is offered by about 2 percent. Ratings of previous e-books of the same author have a similar effect: In column (4), the coefficient on a rating of a previous e-book is positive and statistically significant (p-value < 0.05), indicating that having one additional "star" in rating for a previous e-book increases the probability that a sample is offered by 0.4 percent. Overall, this is evidence against proposition 2. An explanation might be that a rating is an imperfect measure of quality. While on average a better rating should signal better quality, a given e-book might still have the incentive to back-up its good rating by offering a sample. This indicates that the situation is closer to proposition 3, the case where quality is unknown and therefore higher quality sellers have the incentive to offer a sample.

	Dependent variable:			
	Sample Offered (Yes=1)			
	OLS		instrumental variable	
	(1)	(2)	(3)	(4)
Log(Price)	−0.0158***	−0.0165***	0.0209***	0.0558***
	(0.0010)	(0.0010)	(0.0016)	(0.0066)
Rating Available (Yes/No) (Smashw.)			1.1080***	9.7660***
			(0.0195)	(0.5774)
Log(1 + N Ratings) (Smashw.)				−6.7090***
				(0.3975)
Average Rating (Smashw.)			0.0475***	0.3436***
			(0.0041)	(0.0240)
Time Since Published (Weeks)	0.0007***	0.0007***		
	(0.00001)	(0.00001)		
Time Since Published X Average Rating		0.0001***		
		(0.00001)		
Time Since Published X Log(1 + N Ratings)		−0.0001***		
		(0.00001)		
Constant	0.6428***	0.6414***	0.6580***	0.6470***
	(0.0091)	(0.0091)	(0.0132)	(0.0445)
Category (Categorical)	Yes	Yes	Yes	Yes
Subcategory (Categorical)	Yes	Yes	Yes	Yes
Language (Categorical)	Yes	Yes	Yes	Yes
Gender Author	Yes	Yes	Yes	Yes
Length E-Book in Words	Yes	Yes	Yes	Yes
Number Previous Books	Yes	Yes	Yes	Yes
Previous e-Books Available	Yes	Yes	Yes	Yes
Number Ratings Previous Books	Yes	Yes	Yes	Yes
Previous Ratings Available	Yes	Yes	Yes	Yes
Average Ratings Previous E-books	Yes	Yes	Yes	Yes
Number Books	No	No	No	No
Observations	246,157	246,157	246,157	246,157
Note:				*p<0.1; **p<0.05; ***p<0.01 Standard errors in parantheses

Table 8: *Instrumental variables regressions using the number of weeks an e-book is has been on the market as an instrument for whether it has been rated. The dependent variable is whether a given e-book offers a sample.*

Instrumenting for Rating Availability

In this subsection, I present results of an instrumental variables regression using the number of weeks that an e-book has been on sale on *Smashwords* as an instrument for the availability of ratings. This approach is similar to the approach in section 4.1. Table (7) shows the results of the reduced form regressions and of two instrumental variables regressions. The first stage regressions are delegated to table (15) in the appendix. In column (3), I only instrument for the availability of ratings and the interaction between the availability of ratings and the average rating of an e-book. In column (4), I also instrument for the interaction between the availability of ratings and the number of ratings an e-book has received. As is common when using instrumental variables with binary outcome variables, I use Ordinary Least Squares (OLS) regressions in place of a less robust non-linear model.

The results of the instrumental variables regressions in table (7) are similar to my previous results: Being rated increases the probability of offering a sample (p-value < 0.01), while e-books with higher quality as measured by their ratings are more likely to offer samples.

Overall, I find evidence against proposition 6, suggesting that in fact authors of unknown quality are less likely to disclose their product characteristics by offering a sample. In section 6, I show how this results can also be explained by incorporating a fraction of naive buyers into the model.

4.3 Pooling and Price Dispersion

According to proposition 5, for unknown quality we should observe pooling of e-books for which a sample is not offered. Such e-books that do not offer a sample should appear more similar in the eyes of buyers, which should be reflected in a lower price dispersion in my data.

Table (9) shows summaries of the variance of the logarithm of price. The sample is split into two groups, e-books for which a sample is offered and e-books for which no sample is offered. On the upper sub-table in table (9), I look at the difference in variance between both groups by splitting the full sample. Counter to the pooling hypothesis, I find that the variance is larger in the group of e-books where no sample is offered (p-value < 0.001 for both the parametric and non-parametric tests). To control for other sources of heterogeneity, in the middle and lower sub-tables in table (9) I compare the variance of residuals from a regression where I control for all observable characteristics of an e-book. Additionally, I split the sample into e-books without a posted rating (middle sub-table) and e-books with a posted rating (lower sub-table). In both cases, the dispersion of residuals is higher for e-books that do not offer a sample. The difference is statistically significant using the parametric test (p-value<0.001), although the

Sample: All books	Variance	N
log(Price) | Sample Offered = Yes	0.466	$218,780$
log(Price) | Sample Offered = No	0.548	$27,377$
Difference	-0.081	
p-value (H_0 : Equal Variance; Parametric Test)	<0.001	
p-value (H_0 : Same Distribution; Non-Parametric Test)	<0.001	
Sample: Books without rating on Smashwords	Variance	N
$Residuals$(Price) | Sample Offered = Yes	0.351	$194,857$
$Residuals$(Price) | Sample Offered = No	0.424	$25,420$
Difference	-0.073	
p-value (H_0 : Equal Variance; Parametric Test)	<0.001	
p-value (H_0 : Same Distribution; Non-Parametric Siegel–Tukey Test)	0.1	
Sample: Books with rating on Smashwords	Variance	N
$Residuals$(Price) | Sample Offered = Yes	0.317	$23,923$
$Residuals$(Price) | Sample Offered = No	0.386	$1,957$
Difference	-0.068	
p-value (H_0 : Equal Variance; Parametric Test)	<0.001	
p-value (H_0 : Same Distribution; Non-Parametric Siegel–Tukey Test)	0.2	

Table 9: *Variance of prices as a measure of price dispersion. Tests for a statistical significant difference in variance are a F-test based on normality and a non-parametric Siegel–Tukey Rank Sum test adjusted for difference in medians between both groups.*

non-parametric test does not reject the null-hypothesis that both samples are drawn from the same distribution (p-value=0.1 and p-value=0.2). This might, however, be a result of the known low power of the non-parametric test.

To sum up, I do not find any evidence for pooling of e-books for which no sample is observed. To the contrary, I find that price-dispersion within the sub-sample of e-books without a sample is higher than in the sub-sample of e-books where a sample is offered. A possible explanation is that buyers do not take not offering a sample as a bad signal and therefore no pooling occurs. If additionally some authors want to inflate their prices above their quality, they should simultaneously hide their quality by not offering a sample. I show in section (6) how such a reasoning can be incorporated into the theoretical model.

5 Robustness

In this section, I discuss some results regarding the robustness of my results. The detailed tables are delegated to the appendix.

5.1 Using Other Sources of Ratings

Additional to ratings posted on *Smashwords,* I have also collected ratings from the social reading community *Goodreads.com* and from *Amazon.com.* For 15,817 e-books I was able to find ratings on *Goodreads* and for 31,106 e-books I was able to find ratings on *Amazon*. In this section, I show that using these alternative sources as a measure for the availability of information on the quality of e-books produces similar results. As can be seen from the histograms in figure (5) in the appendix, an advantage using these alternative sources is that average ratings on both other platforms are more evenly distributed than on *Smashwords*.

Table (10) in the appendix shows linear regression results replicating the results from section 4.1. Using both alternative sources of ratings as measures of the availability of ratings produces qualitatively the same results as in my initial analysis: The sizes of the coefficients are similar and all effects remain statistically significant (p-value < 0.01).

Table 11 in the appendix shows logit regression results replicating the results from section 4.2. As before, using ratings from *Goodreads* and from *Amazon* as alternative measures for the availability of information on an e-book's quality produces qualitatively the same results, results similar in size and statistical significance. The only exception is that having a good average rating on *Amazon* does not have a statistically significant influence on whether a sample is offered for an e-book on *Smashwords*.

5.2 Exclusion of Outliers

I excluded in my main analysis e-books with a price above $50 and e-books of authors with more than 50 e-books. In this subsection, I show how robust more results are to these exclusions.

In table (12), I show results from a regression with an e-book's price as the dependent variable where I modify the sample size of my dataset. In column (1), I include all e-books on which I have gathered data and in column (2) I further restrict the dataset to only include e-books with a price below $10. All of my results are unaffected, with the exception of the coefficient on whether a sample is offered and whether a previous e-book of the same author is available on *Smashwords* in column (1) where I include all e-books.

In table (13) in the appendix, I show robustness results of the logistic regression for the probability of offering a sample used in section 4.2. In column (1), I use the whole sample collected from *Smashwords*, excluding only e-books with a zero price. Two of my results from section 4.2 are affected by this. The first result is that the coefficient on the average rating of previous e-books by the same author changes from positive to negative. I can show, however, that this reversal is driven by not excluding authors with a disproportionally high number of e-books. In column (2), I exclude one single author who has released 17,000 e-books containing sheet music with musical scores. In column (3), I exclude two additional authors that have released 2,400 and 1,400 e-books each. Making these exclusions, the coefficient changes from negative to positive. This indicates that it is important to exclude authors with too many e-books that are otherwise overrepresented in my sample.

The second result that is affected is the positive effect of being rated on the probability of offering a sample. Although its size stays similar, it is no longer statistically significant.

In column (4), I exclude all e-books with a price larger than \$10. This does not impact my results.

6 Incorporating Naive Consumers into the Model

In this section, I present an extension of the baseline model from section 2. My goal is to explain the two empirical findings that are in conflict with my baseline model of voluntary disclosure: Higher prices of non-rated e-books not offering a sample, and a larger fraction of e-books not offering a sample in the case when e-books are not rated. To explain these results, I extend the baseline model by assuming that a fraction of buyers are naive and fail to anticipate that low-quality sellers are less likely to offer samples. This behavioral assumption is in line with evidence showing that buyers often fail to account for adverse selection or moral hazard in markets with asymmetric information. The consequences of such behavior have been discussed more generally in recent theoretical work incorporating boundedly rational consumers into models of markets with asymmetric information (Eyster and Rabin, 2005; Esponda, 2008). In these models, naivete is modeled as agents failing to adjust their beliefs to the equilibrium behavior of other agents. Such naivete has important implications for contexts such as auctions, credit-card markets, financial services or the insurance industry. That consumer naivete plays also a prominent role in games of voluntary disclosure is suggested by experimental results (Jin et al., 2014; Benndorf et al., 2015). There is also experimental evidence that suggests that sellers take advantage of buyers not accounting for moral hazard in markets with asymmetric information (Henze et al., 2015).

The following reasons suggest that self-publishing and *Smashwords* may be prone to consumer naivete and adverse selection: Self-publishing is a very recent phenomenon with which buyers have limited experience. In traditional publishing, publishers act as gate-keepers by limiting the amount of newly published books. They select the most promising book projects in terms of commercial, artistic, or intellectual potential. Selected book projects typically go through an extended editing process. Although it is a matter of debate how good publishers are in selecting the best books,[27] the average quality of traditionally published books should still be higher than the average quality of self-published books as the entry cost into self-publishing is virtually zero, no preselection takes place, and it is up to the author to employ external editing services. Additionally, while there is a trend of increasing professionalization, the vast majority of self-published authors can be regarded as amateurs. Taken together, it is likely that at least some buyers are unaware of these factors when choosing to buy a self-published e-book. Regarding the signal of not offering a sample, inexperienced buyers do not necessarily have to be aware of authors on *Smashwords* being able to choose whether to offer a sample. On other platforms (e.g. *Amazon*) it is usual that the platform mandates that a sample has to be offered. On *Smashwords*, buyers either need to visit the websites of multiple e-books or to read *Smashwords's* Terms of Conditions to find out about this option for authors.

Extending the Model

The difference to the set-up of the baseline model in section 2 is that I incorporate a fraction $\chi \in (0, 1)$ of buyers who are "naive" in the sense of not adjusting their beliefs to sellers' equilibrium behavior. Hence, they do not anticipate that the expected quality of a seller not offering a sample is lower than the quality of a seller offering a sample. The remaining $1 - \chi$ buyers are "sophisticated" and use Bayesian reasoning to adjust

[27]The book industry is prone to the usual demand uncertainty of creative industries captured in the famous "nobody knows anything" principle coined by the screenwriter William Goldman (Goldman, 1989; Caves, 2000; Canoy et al., 2006).

their beliefs to equilibrium behavior of sellers.[28] Modeling consumer naivete in this way has been introduced to the literature by Eyster and Rabin (2005) and Esponda (2008).

The distinction between naive and sophisticated buyers is only important in the case of unknown quality. When a seller's quality is known, sophisticated reasoning is not needed to infer her quality. The previous results from the baseline model for the case of known quality are therefore unaffected by the presence of naive buyers.

The belief of a naive buyer about the quality of a seller who does not offer a sample is denoted with $E(q) = \theta$. This expectation is not conditioned on which types of sellers choose in equilibrium to offer a sample. A realistic choice would be the average quality of sellers in the market, i.e. $\theta = \frac{3}{2}\bar{\epsilon}$. However, as in the context of self-publishing buyers might also fail to anticipate that worse quality authors select into self-publishing, it is possible that $\theta > \frac{3}{2}\bar{\epsilon}$. The belief of a sophisticated buyer of the quality of a seller not offering a sample is denoted with $E(q|S = n)$. It is conditioned on lower quality types of sellers being more likely to not offer a sample. As in the baseline model, the conditional expectation is denoted with $E(q|S = n) = \frac{1}{2}q^*$, where q^* denotes the quality of the highest type not offering a sample. To limit the cases I have to consider, I assume $\theta > E(q|S = n)$ and that sellers with a quality $q < q^*$ do not offer a sample. Later, I confirm that both conditions hold in equilibrium.

In the case when a seller offers a sample for her product, both naive and sophisticated buyers infer the quality from inspecting the sample. Therefore, it does not matter whether a given buyer is naive or sophisticated.[29] A seller's profit when offering a sample is given by the same expression as in the baseline model:

$$\Pi_s^* = \frac{(\bar{\epsilon} + q)^2}{8\bar{\epsilon}}.$$

The demand function of a seller not offering a sample now differs from the baseline

[28] Alternatively, χ can be thought of as a parameter governing the degree to which buyers are naive, or "cursed" in the terminology of Eyster and Rabin (2005).

[29] The Behavioral IO literature terms buyers who adjust their naive beliefs to such direct information as naive but educable (Gabaix and Laibson, 2006; Heidhues et al., 2014; Murooka, 2015).

model. It now includes the option to sell at a price where only naive buyers demand the product:

$$D_n(p_n) = \begin{cases} 0 & \text{for } p_n > \theta \\ \chi & \text{for } E(q|S=n) < p_n \leq \theta \\ 1 & \text{for } p_n \leq E(q|S=n). \end{cases}$$

Given that a seller should either set a price of $p_n = \theta$ if she targets only naive buyers, or set a price of $p_n = E(q|S = n)$ if she targets all buyers, profits when not offering a sample are given by

$$\Pi_n^* = \begin{cases} \chi\theta & \text{for } p_n = \theta \\ E(q|S=n) & \text{for } p_n = E(q|S=n). \end{cases}$$

Note that only if $\chi\theta > E(q|S = n)$ sellers will find it profitable to target naive buyers. This will be the case only if there are sufficiently many naive buyers (χ sufficiently large) and their beliefs are sufficiently over-optimistic (θ sufficiently larger than $E(q|S = n)$). Whether this condition is fulfilled does not depend on a seller's true type. Either every seller not offering a sample will target only naive buyers or every seller not offering a sample will target both types of buyers. As only in the case where $\chi\theta > E(q|S = n)$ the presence of naive buyers has an influence on sellers' equilibrium behavior, in the following I focus on the case where this condition holds. Afterwards, I show under which parameter values the condition is fulfilled.

The types of sellers who offer a sample is determined as in the baseline model by the seller of type q^* who is indifferent between offering and not offering a sample. The indifferent seller is found by solving

$$\frac{(\overline{\epsilon} + q^*)^2}{8\overline{\epsilon}} = \chi\theta$$

for q^*, which gives the solution

$$q^* = \sqrt{8\chi\theta\bar{\epsilon}} - \bar{\epsilon}. \tag{10}$$

All sellers with $q > q^*$ offer a sample and all sellers with $q < q^*$ do not offer a sample. The beliefs of sophisticated buyers are given by

$$E(q|S = n) = \sqrt{2\chi\theta\bar{\epsilon}} - \frac{\bar{\epsilon}}{2}. \tag{11}$$

Inserting equation (11) into the condition $\chi\theta > E(q|S = n)$, which guarantees that sellers not offering a sample find it profitable to target naive buyers, gives the condition

$$\chi\theta > \frac{\bar{\epsilon}}{2}. \tag{12}$$

This condition also guarantees that the weaker initial condition $\theta > E(q|S = n)$ is always fulfilled. We can use condition (12) to compare the fraction of sellers not offering a sample between the case where some buyers are naive, given by equation (10), with the case where all buyers are sophisticated, given by equation (9). From this comparison, we see that if (12) is fulfilled, the rate of sellers not offering a sample is higher when some buyers are naive. Naive buyers thus exert an externality on the market that leads to less sellers offering a sample. On the other hand, sophisticated buyers also exert an externality by disciplining more low-quality sellers to offer a sample.

Explaining Higher Prices of Non-Disclosing Sellers

In my empirical analysis, I find that non-rated e-books with no sample have a higher average price than e-books where authors offer a sample. This result is in conflict with proposition 4 of the baseline model. In the context of my extended model, I explain this result by showing for which parameter values $E(p_n^*) > E(p_s^*)$. After inserting equilibrium prices, this gives the condition:

$$\chi\theta > E\left(\frac{\overline{\epsilon}+q}{2}\Big|q > q^*\right).$$

By using the assumption that the highest type is $\bar{q} = 3\bar{\epsilon}$ and that the distribution of feasible qualities is uniform, combined with the lowest type offering a sample q^* being given by equation (10), we arrive at the following condition

$$\chi\theta > 2\bar{\epsilon}. \tag{13}$$

The following proposition summarizes this result:

Proposition 7. *When the quality of products is unknown to buyers, if both the number of naive buyers and their beliefs of the average quality of sellers are sufficiently high, offering a sample is associated with a lower price.*

Although sellers of a lower quality who do not offer a sample set higher prices, their profits are lower than the profits of high-quality sellers offering a sample. The reason is that lower quality sellers only sell to naive buyers while higher quality sellers sell to both naive and sophisticated buyers.

Explaining Less Disclosure When Quality Unknown

In my empirical analysis, I find that the probability that a sample is offered is lower for unrated e-books than for rated e-books. This result is in conflict with proposition 6 of the baseline model. I show how the extended model can account for this result by assuming that a sufficient number of buyers are naive with sufficiently high expectations of sellers' quality. The quantity of sellers who offer a sample in the case of known quality is given by

$$3\bar{\epsilon} - q^* = 3\bar{\epsilon} - \left(\sqrt{8\chi\theta\bar{\epsilon}} - \bar{\epsilon}\right). \tag{14}$$

The quantity of sellers offering a sample in the case of known quality is given by equation (6). We have to find parameter values for which (14) is smaller than (6). The condition for which (14) is smaller than (6) is given by:

$$\chi\theta > \frac{(9+4\sqrt{2})\bar{\epsilon}}{8} \approx 1.83\bar{\epsilon}. \tag{15}$$

Notice that when condition (13) is fulfilled, condition (15) is also fulfilled. The following proposition summarizes this result:

Proposition 8. *When both the number of naive buyers and their beliefs of the average quality of sellers are sufficiently high, fewer sellers offer a sample when quality is unknown as compared to when quality is known.*

7 Conclusion

In this paper, I have empirically examined the link between disclosure of product information and price in a setting where I observe both sellers of known and unknown quality, as indicated whether posted online ratings are available for their products. While in the case of unknown quality, I find a negative relationship between price and whether a seller discloses, for known quality I observe a positive relationship. The positive relationship in the case of known quality confirms the prediction of theory in the form of sellers earning a "match-premium" when disclosing horizontal information (Johnson and Myatt, 2006; Bar-Isaac et al., 2010; Sun, 2011; Celik, 2014). The negative relationship in the case of unknown quality is the opposite of what classical unraveling theory predicts (Grossman, 1981; Milgrom, 1981; Jovanovic, 1982; Viscusi, 1978) predicts. This can, however, be explained by buyers not interpreting not disclosing as a bad signal of a seller's quality and sellers of a low quality trying to take advantage of such buyers by not disclosing. This interpretation is consistent with similar findings from the lab by Jin et al. (2014) and Henze et al. (2015). I show how a model of voluntary disclosure needs to be adapted to account for these results by including a fraction of naive buyers who are inattentive of the bad signal that non-disclosure sends. This links the literature on voluntary disclosure to the recently emerging literature on behavioral industrial organization.[30]

In terms of policy, my results have the following implications: The possibility that sellers of unknown quality do not reveal their low quality implies that disclosure policies are especially important in markets where there is a large fraction of such sellers, e.g. in emerging markets or markets with significant entry of new sellers due to low entry costs. Even if there are credible and non-costly ways for sellers to voluntarily disclose information about their products, as is the case with digital samples, disclosure policies might be necessary as buyers are inattentive to the signal of not disclosing. Furthermore, such consumer inattentiveness could possibly trigger even further entry by low-quality sellers, amplifying problems with asymmetric information and adverse selection, as it becomes harder for high-quality sellers to enter buyers' consideration sets. Overall, this suggests that a mandatory disclosure policy should considerably improve market outcomes.

The fact that in the case of known quality sellers, I observe that disclosure is linked to higher price, suggest that sellers take advantage of a better match due to disclosure by increasing their price. This positive effect needs to be taken into account when deciding whether to implement mandatory disclosure policies, although even net of the price effect the better match between product characteristics and consumers' taste should still improve efficiency by improving the allocation goods.

[30]See Ellison (2006) and Spiegler (2011) for surveys of the field.

Appendix

Figure 4: *Example of an e-book's web-page on* Smashwords.com.

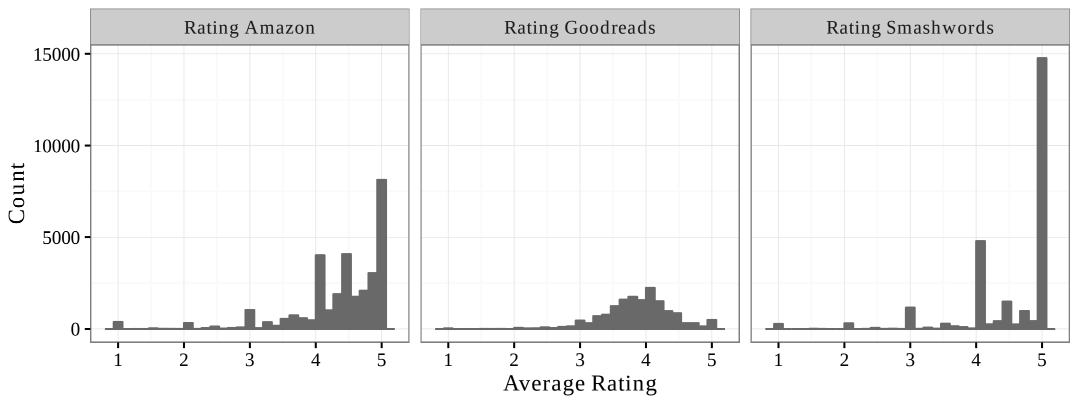

Figure 5: *Histogram of average posted ratings on* Amazon, Goodreads *and* Smashwords.

	Dependent variable:	
	Log(Price)	
	(1)	(2)
Sample Offered (Yes/No)	-0.103^{***}	-0.102^{***}
	(0.007)	(0.007)
Log(1 + N Ratings) (Goodreads)	0.019^{***}	
	(0.002)	
Rating Available (Yes/No) (Goodreads)	-0.128^{***}	
	(0.019)	
Average Rating (Goodreads)	0.051^{***}	
	(0.008)	
Log(1 + N Ratings) (Amazon)		0.014^{***}
		(0.003)
Rating Available (Yes/No) (Amazon)		-0.035^{***}
		(0.013)
Average Rating (Amazon)		0.026^{***}
		(0.005)
Previous Book Available (Yes/No)	-0.119^{***}	-0.119^{***}
	(0.010)	(0.010)
Rating Available Previous Book (Smashw.)	-0.218^{***}	-0.220^{***}
	(0.010)	(0.010)
Sample Offered X Rating Available (Goodreads)	0.093^{***}	
	(0.015)	
Sample Offered X Rating Available (Amazon)		0.035^{***}
		(0.012)
Sample Offered X Previous Book Available	0.034^{***}	0.034^{***}
	(0.010)	(0.010)
Sample Offered X Rating Available Previous Book	0.082^{***}	0.085^{***}
	(0.011)	(0.011)
Constant	0.774^{***}	0.765^{***}
	(0.071)	(0.071)
Category (Categorical)	Yes	Yes
Subcategory (Categorical)	Yes	Yes
Language (Categorical)	Yes	Yes
Gender Author (Categorical)	Yes	Yes
Year Published (Categorical)	Yes	Yes
Length E-Book in Words	Yes	Yes
Time Since Published (Weeks)	Yes	Yes
Number Previous Books	Yes	Yes
Number Books	Yes	Yes
Observations	246,157	246,089
R^2	0.253	0.253
Note:		*p<0.1; **p<0.05; ***p<0.01
		Robust standard errors in parantheses

Table 10: *Linear regressions with an e-book's price as the dependent variable. Ratings from Goodreads.com and Amazon.com are used as alternative measures for the availability of information of an e-book's quality.*

	Dependent variable:	
	Sample Offered (Yes=1)	
	(1)	(2)
Log(Price)	-0.012^{***}	-0.012^{***}
	(0.001)	(0.001)
Rating Available (Yes/No) (Goodreads)	0.013^{**}	
	(0.006)	
Log(1 + N Ratings) (Goodreads)	-0.010^{***}	
	(0.001)	
Average Rating (Goodreads)	0.015^{***}	
	(0.004)	
Rating Available (Yes/No) (Amazon)		0.023^{***}
		(0.003)
Log(1 + N Ratings) (Amazon)		-0.015^{***}
		(0.001)
Average Rating (Amazon)		0.001
		(0.002)
Previous E-books Available (Yes/No)	0.003^{*}	0.003
	(0.002)	(0.002)
Previous Ratings Available (Yes/No)	0.011^{***}	0.011^{***}
	(0.001)	(0.001)
Average Rating Previous E-books	0.005^{***}	0.005^{***}
	(0.002)	(0.002)
Constant	0.203^{**}	0.204^{**}
	(0.082)	(0.082)
Category (Categorical)	Yes	Yes
Subcategory (Categorical)	Yes	Yes
Language (Categorical)	Yes	Yes
Gender Author (Categorical)	Yes	Yes
Year Published (Categorical)	Yes	Yes
Length E-Book in Words	Yes	Yes
Time Since Published (Weeks)	Yes	Yes
Number Previous Books	Yes	Yes
Observations	246,157	246,089
Note:	*p<0.1; **p<0.05; ***p<0.01 Coefficients give marginal effects estimates; standard errors in paranthenses	

Table 11: *Logit regressions with the probability that a sample is offered for an e-book as the dependent variable. Ratings from Goodreads.com and Amazon.com are used as alternative measures for the availability of information on an e-book's quality.*

	Dependent variable:	
	Log(Price)	
	(1)	(2)
Sample Offered (Yes/No)	-0.103^{***}	-0.102^{***}
	(0.007)	(0.007)
Log(1 + N Ratings) (Goodreads)	0.019^{***}	
	(0.002)	
Rating Available (Yes/No) (Goodreads)	-0.128^{***}	
	(0.019)	
Average Rating (Goodreads)	0.051^{***}	
	(0.008)	
Log(1 + N Ratings) (Amazon)		0.014^{***}
		(0.003)
Rating Available (Yes/No) (Amazon)		-0.035^{***}
		(0.013)
Average Rating (Amazon)		0.026^{***}
		(0.005)
Previous Book Available (Yes/No)	-0.119^{***}	-0.119^{***}
	(0.010)	(0.010)
Rating Available Previous Book (Smashw.)	-0.218^{***}	-0.220^{***}
	(0.010)	(0.010)
Sample Offered X Rating Available (Goodreads)	0.093^{***}	
	(0.015)	
Sample Offered X Rating Available (Amazon)		0.035^{***}
		(0.012)
Sample Offered X Previous Book Available	0.034^{***}	0.034^{***}
	(0.010)	(0.010)
Sample Offered X Rating Available Previous Book	0.082^{***}	0.085^{***}
	(0.011)	(0.011)
Constant	0.774^{***}	0.765^{***}
	(0.071)	(0.071)
Category (Categorical)	Yes	Yes
Subcategory (Categorical)	Yes	Yes
Language (Categorical)	Yes	Yes
Gender Author (Categorical)	Yes	Yes
Year Published (Categorical)	Yes	Yes
Length E-Book in Words	Yes	Yes
Time Since Published (Weeks)	Yes	Yes
Number Previous Books	Yes	Yes
Number Books	Yes	Yes
Observations	246,157	246,089
R^2	0.253	0.253
Note:	*p<0.1; **p<0.05; ***p<0.01	
	Robust standard errors in parantheses	

Table 12: *Additional linear regression with price as a dependent variable as robustness checks. In column (1) I include all e-books on which I have gathered data. In column (2) all e-books with a price above or equal to \$10 are excluded.*

	Dependent variable:			
	Sample Offered (Yes=1)			
	All E-books	Excl. Sheet Music	Excl. Authors More 1000 e-B.	Price < 10 Dol.
	(1)	(2)	(3)	(4)
Log(Price)	-0.015^{***}	-0.018^{***}	-0.017^{***}	-0.010^{***}
	(0.001)	(0.001)	(0.001)	(0.001)
Rating Available (Yes/No) (Smashw.)	0.004	0.004	0.006	0.006
	(0.004)	(0.004)	(0.004)	(0.004)
Log(1 + N Ratings) (Smashw.)	-0.011^{***}	-0.012^{***}	-0.011^{***}	-0.011^{***}
	(0.003)	(0.004)	(0.004)	(0.004)
Average Rating (Smashw.)	0.022^{***}	0.023^{***}	0.022^{***}	0.021^{***}
	(0.002)	(0.002)	(0.002)	(0.002)
Previous E-books Available (Yes/No)	0.011^{***}	0.019^{***}	0.014^{***}	0.003^{*}
	(0.001)	(0.002)	(0.002)	(0.002)
Previous Ratings Available (Yes/No)	0.018^{***}	0.018^{***}	0.006^{***}	0.009^{***}
	(0.001)	(0.001)	(0.001)	(0.002)
Average Rating Previous E-books	-0.007^{***}	-0.004^{***}	0.002^{**}	0.004^{***}
	(0.001)	(0.001)	(0.001)	(0.002)
Constant	0.224^{***}	0.243^{***}	0.246^{***}	0.207^{**}
	(0.076)	(0.084)	(0.083)	(0.082)
Category (Categorical)	Yes	Yes	Yes	Yes
Subcategory (Categorical)	Yes	Yes	Yes	Yes
Language (Categorical)	Yes	Yes	Yes	Yes
Gender Author (Categorical)	Yes	Yes	Yes	Yes
Year Published (Categorical)	Yes	Yes	Yes	Yes
Length E-Book in Words	Yes	Yes	Yes	Yes
Time Since Published (Weeks)	Yes	Yes	Yes	Yes
Number Previous Books	Yes	Yes	Yes	Yes
Observations	311,582	294,343	290,471	242,377

Note: *p<0.1; **p<0.05; ***p<0.01
Coefficients give marginal effects estimates;
standard errors in paranthenses

Table 13: *Additional logistic regressions as robustness checks with probability of offering a sample as a dependent variable. In column (1) I exclude all e-books with a price above or equal to \$10 and in column (2) all e-books with a price above or equal \$5. In column (3) I include the e-books by one particular seller who has published 17.000 sheet music (musical scores) on* Smashwords.

	Dependent variable:			
	Rati.Avail.	(Rat.Avail. X Sample)	(Rat.Avail. X Av.Rat.)	[Rat.Avail. X Log(1 + N Rat.)]
	First Stage			
	(1)	(2)	(3)	(4)
Sample Offered (Yes/No)	-0.0041^{**}	0.0226^{***}	0.0018^{***}	-0.0034^{***}
	(0.0019)	(0.0021)	(0.0006)	(0.0009)
Time Since Published (Weeks)	0.0005^{***}			
	(0.00001)			
Time Since Published X Sample Offered		0.0005^{***}		
		(0.00001)		
Time Since Published X Average Rating			0.0051^{***}	
			(0.000004)	
Time Since Published X Log(1 + N Ratings)				0.0048^{***}
				(0.000004)
Constant	-0.0518^{***}	-0.0680^{***}	0.0010	0.0452^{***}
	(0.0086)	(0.0083)	(0.0027)	(0.0041)
Category (Categorical)	Yes	Yes	Yes	Yes
Subcategory (Categorical)	Yes	Yes	Yes	Yes
Language (Categorical)	Yes	Yes	Yes	Yes
Gender Author	Yes	Yes	Yes	Yes
Length E-Book in Words	Yes	Yes	Yes	Yes
Number Previous Books	Yes	Yes	Yes	Yes
Number Ratings Previous Books	Yes	Yes	Yes	Yes
Number Books	Yes	Yes	Yes	Yes
Observations	246,157	246,157	246,157	246,157
R^2	0.1353	0.1367	0.8545	0.8598
Note:				*p<0.1; **p<0.05; ***p<0.01 Robust standard errors in parantheses

Table 14: *First stage regressions used in instrumental variables regressions in table 6.*

	Dependent variable:		
	Rati.Avail.	(Rat.Avail. X Av.Rat.)	[Rat.Avail. X Log(1 + N Rat.)]
	(1)	(2)	(3)
Time Since Published (Weeks)	0.0007^{***}		
	(0.00001)		
Time Since Published X Average Rating		0.0051^{***}	
		(0.000004)	
Time Since Published X Log(1 + N Ratings)			0.0048^{***}
			(0.000004)
Log(Price)	-0.0334^{***}	-0.0001	-0.0068^{***}
	(0.0010)	(0.0003)	(0.0005)
Constant	-0.0138	0.0017	0.0320^{***}
	(0.0086)	(0.0027)	(0.0041)
Category (Categorical)	Yes	Yes	Yes
Subcategory (Categorical)	Yes	Yes	Yes
Language (Categorical)	Yes	Yes	Yes
Gender Author	Yes	Yes	Yes
Length E-Book in Words	Yes	Yes	Yes
Number Previous Books	Yes	Yes	Yes
Previous e-Books Available	Yes	Yes	Yes
Number Ratings Previous Books	Yes	Yes	Yes
Previous Ratings Available	Yes	Yes	Yes
Average Ratings Previous E-books	Yes	Yes	Yes
Number Books	No	No	No
Observations	246,157	246,157	246,157
R^2	0.1173	0.8547	0.8593
Note:			*p<0.1; **p<0.05; ***p<0.01 Standard errors in parantheses

Table 15: *First stage regressions used in instrumental variables regressions in table 8.*

	1:	2:	3:	4:	5:	6:	7:	8:	9:	10:	11:	12:	13:	14:	15:	16:	17:
1: Log(Price)	1																
2: Sample Offered (Yes/No)	−0.06	1															
3: Sample Size (Percent)	−0.15	0.61	1														
4: Rating Available (Yes/No) (Smashw.)	−0.08	0.04	0.07	1													
5: Rating Available (Yes/No) (Goodreads)	0.04	0.01	−0.01	0.13	1												
6: Rating Available (Yes/No) (Amazon)	0.04	0.01	−0.01	0.07	0.14	1											
7: Average Rating (Smashw.)	0.08	0.03	−0.01		−0.01	0.04	1										
8: Average Rating (Goodreads)	0.09	−0.0003	−0.01	0.04		0.02	0.24	1									
9: Average Rating (Amazon)	0.07	−0.01	0.002	0.03	−0.02		0.27	0.44	1								
10: Log(1 + N Ratings) (Smashw.)	−0.07	0.04	0.07	0.87	0.15	0.07	0.06	0.04	0.02	1							
11: Log(1 + N Ratings) (Goodreads)	0.04	0.01	−0.02	0.14	0.93	0.14	0.003	0.09	−0.01	0.17	1						
12: Log(1 + N Ratings) (Amazon)	0.05	0.001	−0.02	0.10	0.21	0.84	0.04	0.05	0.08	0.11	0.24	1					
13: Log(Length E-Book in Words)	0.38	0.04	−0.08	0.09	0.16	0.16	0.11	0.17	0.17	0.10	0.17	0.17	1				
14: Log(1 + Time Since Published in Weeks)	−0.02	0.15	0.16	0.15	0.09	0.06	−0.10	−0.12	−0.12	0.16	0.08	0.04	−0.03	1			
15: Log(1 + Number Previous Books)	−0.03	−0.02	−0.06	−0.11	−0.06	−0.09	−0.10	−0.04	−0.14	−0.12	−0.06	−0.08	−0.20	−0.12	1		
16: Log(1 + Number Ratings Previous Books)	−0.11	0.06	0.08	0.20	0.08	0.04	0.04	0.09	0.01	0.20	0.09	0.07	0.01	0.01	0.44	1	
17: Log(Number Books)	−0.06	−0.004	−0.03	−0.06	−0.03	−0.08	−0.13	−0.10	−0.19	−0.07	−0.02	−0.07	−0.23	0.02	0.91	0.39	1

Table 16: *Correlation Matrix*

References

Aguzzoni, L., Argentesi, E., Ciari, L., Duso, T., and Tognoni, M. (Forthcoming). "Ex-post Merger Evaluation in the UK Retail Market for Books." *Journal of Industrial Economics*.

Bagwell, K. (2007). "The Economic Analysis of Advertising." In R. Schmalensee (Ed.), *Handbook of Industrial Organization*, vol. 3, 1701–1844, Elsevier.

Bar-Isaac, H., Caruana, G., and Cuñat, V. (2010). "Information Gathering and Marketing." *Journal of Economics and Management Strategy*, *19*(2), 375–401.

Beck, J. (2007). "The Sales Effect of Word of Mouth: A Model for Creative Goods and Estimates for Novels." *Journal of Cultural Economics*, *31*(1), 5–23.

Bederson, B., Jin, G. Z., Leslie, P., Quinn, A., and Zou, B. (2015). "Incomplete Disclosure: Evidence of Signaling and Countersignaling." *Working Paper*.

Benndorf, V., Kübler, D., and Normann, H.-T. (2015). "Privacy Concerns, Voluntary Disclosure of Information, and Unraveling: An Experiment." *European Economic Review*, *75*, 43–59.

Berry, S., Levinsohn, J., and Pakes, A. (1995). "Automobile Prices in Market Equilibrium." *Econometrica*, 841–890.

Bowker (2013). "Self-Publishing in the United States, 2007-2012." *http://www.bowker.com/assets/downloads/products/selfpublishingpubcounts_2007_2012.pdf*.

Bowker (2014a). "Self-Publishing Continues to Grow in U.S., Says Bowker." *www.bowker.de, 8.10.2015*.

Bowker (2014b). "Self-Publishing in the United States, 2007-2013." *http://media.bowker.com/documents/bowker_selfpublishing_report2013.pdf*.

Bun, M. J., and Harrison, T. D. (2014). "OLS and IV Estimation of Regression Models Including Endogenous Interaction Terms." *UVA Amsterdam Discussion Paper: 2014/02*.

Cabral, L., and Hortacsu, A. (2010). "The Dynamics of Seller Reputation: Evidence from eBay." *Journal of Industrial Economics*, *58*(1), 54–78.

Canoy, M., van Ours, J., and Van Der Ploeg, F. (2006). "The Economics of Books." In V. A. Ginsburgh, and D. Throsby (Eds.), *Handbook of the Economics of Art and Culture*, vol. 1, 721–761, Elsevier.

Caves, R. E. (2000). *Creative Industries: Contracts Between Art and Commerce*. Harvard University Press.

Celik, L. (2014). "Information Unraveling Revisited: Disclosure of Horizontal Attributes." *Journal of Industrial Economics*, *62*(1), 113–136.

Chamberlin, E. (1933). *The Theory of Monopolistic Competition*. Cambridge: Harvard University Press.

Coker, M. (2012). "How to Sell EBooks at the Apple iBookstore." *Smashwords Blog available at http://blog.smashwords.com/2012/11/how-to-sell-ebooks-at-apple-ibookstore.html, 3.11.2012*.

Colao, J. (2012). "Apple's Biggest (Unknown) Supplier of E-Books." *Forbes.com available at http://www.forbes.com/sites/jjcolao/2012/06/07/apples-biggest-unknown-supplier-of-e-books, 7.6.2012*.

De los Santos, Babur, and Wildenbeest, M. R. (2015). "E-book Pricing and Vertical Restraints." *Working Paper*.

Dranove, D., and Jin, G. Z. (2010). "Quality Disclosure and Certification: Theory and Practice." *Journal of Economic Literature*, *48*(4), 935–963.

Ellison, G. (2006). "Bounded Rationality in Industrial Organization." In R. Blundell (Ed.), *Advances in Economics and Econometrics*, vol. 2, 142–174.

Esponda, I. (2008). "Behavioral Equilibrium in Economies with Adverse Selection." *American Economic Review*, *98*(4), 1269–1291.

Eyster, E., and Rabin, M. (2005). "Cursed Equilibrium." *Econometrica*, *73*(5), 1623–1672.

Flood, A. (2015). "Many Authors' Earnings Fall Below $500pa, Survey Finds." *The Guardian http://www.theguardian.com/books/2015/jan/23/authors-earnings-fall-350pa-survey, 23.1.2015*.

Gabaix, X., and Laibson, D. (2006). "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets." *Quarterly Journal of Economics*, *121*(2), 505–540.

Gelman, A., and Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

Genesove, D., and Simhon, A. (2015). "Seasonality and the Effect of Advertising on Price." *Journal of Industrial Economics*, *63*(1), 199–222.

Goldman, W. (1989). *Adventures in the Screen Trade*. Grand Central Publishing.

Grossman, S. J. (1981). "The Informational Role of Warranties and Private Disclosure about Product Quality." *Journal of Law and Economics*, *24*(3), 461–483.

Heidhues, P., Köszegi, B., and Murooka, T. (2014). "Inferior Products and Profitable Deception." *Working Paper*.

Henze, B., Schuett, F., and Sluijs, J. P. (2015). "Transparency in Markets for Experience Goods: Experimental Evidence." *Economic Inquiry*, *53*(1), 640–659.

Hotz, V., and Xiao, M. (2013). "Strategic Information Disclosure: The Case of Multiattribute Products with Heterogeneous Consumers." *Economic Inquiry*, *51*(1), 865–881.

Ispano, A., and Schwardmann, P. (2016). "Competitive Pricing and Quality Disclosure to Cursed Consumers." *Working Paper available at https://sites.google.com/site/peterschwardmann*.

Janssen, M. C. W., and Teteryatnikova, M. (2015). "Competition and Disclosure of Comparative or Non-Comparative Horizontal Product Information." *Journal Industrial Economics, Forthcoming*.

Jin, G. Z. (2005). "Competition and Disclosure Incentives: An Empirical Study of HMOs." *RAND Journal of Economics*, *36*(1), 93–112.

Jin, G. Z., and Kato, A. (2006). "Price, Quality, and Reputation: Evidence from an Online Field Experiment." *RAND Journal of Economics*, *37*(4), 983–1005.

Jin, G. Z., Luca, M., and Martin, D. (2014). "Is No News (Perceived as) Bad News? An Experimental Investigation of Information Disclosure." *NBER Working Paper No. 21099*.

Johnson, J. P., and Myatt, D. P. (2006). "On the Simple Economics of Advertising, Marketing, and Product Design." *American Economic Review*, *96*(3), 756–784.

Jovanovic, B. (1982). "Truthful Disclosure of Information." *Bell Journal of Economics*, *13*(1), 36–44.

Koessler, F., and Renault, R. (2012). "When Does a Firm Disclose Product Information?" *RAND Journal of Economics*, *43*(4), 630–649.

Luca, M., and Smith, J. (2015). "Strategic disclosure: The case of business school rankings." *Journal of Economic Behavior & Organization*, *112*, 17–25.

Mathios, A. D. (2000). "The Impact of Mandatory Disclosure Laws on Product Choices: An Analysis of the Salad Dressing Market." *Journal of Law and Economics*, *43*(2), 651–678.

Milgrom, P. R. (1981). "Good News and Bad News: Representation Theorems and Applications." *Bell Journal of Economics*, 380–391.

Murooka, T. (2015). "Deception under Competitive Intermediation." *Working Paper*.

Nabokov, V. (2012). *Vladimir Nabokov: Selected Letters 1940-1977*. Boston: Houghton Mifflin Harcourt.

Nelson, P. (1970). "Information and Consumer Behavior." *Journal of Political Economy*, *78*(2), 311–329.

Ozga, S. (1960). "Imperfect Markets through Lack of Knowledge." *Quarterly Journal of Economics*, *74*(1), 29–52.

Senior, A. (2012). "Ebook Sales Are Being Driven by Downmarket Genre Fiction." *The Guardian available at http://www.theguardian.com/media/2012/feb/05/ebook-sales-downmarket-genre, 5.2.2012*.

Spiegler, R. (2011). *Bounded Rationality and Industrial Organization*. Oxford University Press.

Stedman, A. (2015). "Fifty Shades' Spinoff 'Grey' Copy Reportedly Stolen From Publisher." *Variety available at http://variety.com/2015/biz/news/ifty-shades-spinoff-grey-stolen-from-publisher-1201516843, 10.6.2015*.

Stigler, G. J. (1961). "The Economics of Information." *Journal of Political Economy*, (69), 213–225.

Sun, M. (2011). "Disclosing Multiple Product Attributes." *Journal of Economics and Management Strategy*, *20*(1), 195–224.

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

Tadelis, S., and Zettelmeyer, F. (2015). "Information Disclosure as a Matching Mechanism: Theory and Evidence from a Field Experiment." *American Economic Review*, *105*(2), 886–905.

Viscusi, W. (1978). "A Note on Lemons Markets with Quality Certification." *Bell Journal of Economics*, *9*(1), 277–279.

Waldfogel, J. (2012). "Copyright Protection, Technological Change, and the Quality of New Products: Evidence from Recorded Music since Napster." *Journal of Law and Economics*, *55*(4), 715–740.

Waldfogel, J. (2015). "Cinematic Explosion: New Products, Unpredictability, and Realized Quality in the Digital Era." *Journal Industrial Economics*, *Forthcoming*.

Waldfogel, J., and Reimers, I. (2015). "Storming the Gatekeepers: Digital Disintermediation in the Market for Books." *Information Economics and Policy*, *31*, 47–58.

Zindler, P. H. (2009). *Self-Publishing and Marketing from the Trenches*. Innovo Publishing, LLC.