

UC Berkeley

Recent Work

Title

Judging a Book by its Cover — Online Previews and Book Sales

Permalink

<https://escholarship.org/uc/item/69n9j5xs>

Author

Wallenberg, Fredrik

Publication Date

2009-04-01

Judging a Book by its Cover

Online Previews and Book Sales*

Fredrik Wallenberg
School of Information
University of California Berkeley

UC Berkeley School of Information Report 2009-031
April 2009

Abstract

There has been a continued debate about the benefits and cost associated with providing free samples of information goods on the Internet. Some argue that the samples lead to increased sales through increased awareness of the good while others claim that the previews and samples cannibalize sales. In this paper I present a unifying model where we show that information about the good, specifically samples/trial versions/previews etc of the good have both a sales promoting and cannibalizing effect and that either of the two can be dominant. I then set up an experiment in which I look at the impact on the sales of a specific set of books from the enabling of full text search of the contents and previews of pages relevant to the search query. I find no significant impact on sales from these previews. The sample available to me is however on the small side and also from a very specific genre, both of which impact the results.

*I want to thank Coye Cheshire, John Chuang, Hal Varian and Florian Zettelmeyer for invaluable feedback and criticism when finalizing this project and my dissertation. I also want to thank Glenn Woroch for being a sounding board throughout the project and John Morgan and Bronwyn Hall for assistance in designing the experiment. Allen Noren, Tim Allwin, Tim O'Reilly and Roger Magoulas at O'Reilly Media were very helpful in helping me set up the experiment, in the end, we couldn't run it together. Thanks to Jim King at Nielsen Book Scan I got access to sales data for the period of the experiment. Finally I want to thank my wife, Jennifer, for a lot of editing, moral support and coding. The author can be reached at fredrik@ischool.berkeley.edu.

1 Introduction

In brick-and-mortar businesses, product previews are nothing new. A customer can listen to music or thumb through books before having to make any kind of purchase commitment. Long before we buy our movie tickets we have been able to see promotional videos for the movie. When it comes to similar analysis of previews on the internet there is less consensus. The proponents cite the same benefits as have been documented in the physical world for a long time while the critics are concerned about unauthorized copies being circulated on the internet.

In the marketing literature previews can be traced back to [Bauer \(1960\)](#). He first suggested that the consumer decision making (in purchasing) should be seen from the perspective of risk taking. Specifically he notes that there are real and perceived risks associated with purchases such as loss of money (the product wasn't as good as expected), loss of face ("I can't believe you bought that!"), and risk of bodily harm (the product is dangerous). [Cox and Rich \(1964\)](#) found that "not having to see [the items]" was the second most important feature that made housewives buy items over the phone (rather than in a store)—implying that, for some goods, being able to preview the good will increase the likelihood of a purchase. [Roselius \(1971\)](#) reports consumer rankings of eleven different risk reduction "strategies", one of which was *free samples* (which, in many ways, resemble previews the way we think of them). Roselius further defines perceived risk as being associated with four distinct types of losses: Time Loss, Hazard Loss, Ego Loss and Money Loss. Free samples were perceived as equally effective for time and money loss, and significantly more helpful for ego loss and significantly less so for hazard loss. In the context of creative works, hazard loss (the failure of the product is dangerous to our health) seems of limited relevance whereas the other three most certainly apply.

In the digital realm, the "previewing is marketing" argument has been used in the support of file sharing ever since the Napster case (see for example [Oberholzer and Strumpf, 2004](#); [Michel, 2004](#); [BBC News](#); [Borland](#); [Menta](#)). It also surfaces in the "snippets" returned by search engines.¹ The concern over copying (since any transmission over the internet by its very nature creates copies) that were argued back in 2000 keeps surfacing over and over. When taking the previews argument all the way to file-sharing (as opposed to limited samples) we have research such as [Liebowitz \(2002\)](#) that show how file-sharing does indeed

¹It is a widely believed, if not scientifically proven, fact that snippets improve search results by making it easier to determine the usefulness and relevancy of each particular result.

affect sales in a negative way and [Smith and Telang \(2008\)](#) that find that the availability of pirated movies had no impact on DVD sales.

The Authors' Guild's decision to sue Google highlights this issue to some extent.² However, that lawsuit is also about control with the plaintiffs arguing that opt-out is not enough under copyright law. Google responds that their uses are Fair Use. This paper does nothing to address those issues but strive to respond to the question of whether services such as *Google Book Search* and Amazon's *Search Inside!* help or hurt sales of physical (or electronic) books. The Author's Guild claim that *Google Book Search* is for commercial purposes (through the sale of advertising) and that (from paragraph 35) . . .

. . . Google's acts have caused, and unless restrained, will continue to cause damages and irreparable injury to the Named Plaintiffs and the Class through:

- a. continued copyright infringement of the Works and/or the effectuation of new and further infringements;
- b. depreciation in the value and ability to license and sell their Works;
- c. lost profits and/or opportunities; and
- d. damage to their goodwill and reputation.

In general, controversy stems from the fact that viewing digital content over the internet automatically creates a local copy that, in many cases, can be saved, printed or otherwise reused by the consumer without further interaction with the seller (and without actually buying the content). This concern can be addressed in a number of ways. One way is to employ technical limitations on what the user can do through digital rights management systems. Another is to limit the content available in the preview to minimize the possibility that the user can satisfy his need for the work simply by viewing/reading/listening to the preview.

In this research project I analyze the effect on the sale of one type of works by one publisher by previews offered by one content aggregator / retail channel. It is clear that previewing serves different purposes at different stages of the information search and acquisition process, and that not all types of contents will be impacted the same. Within those limitations I hope this paper will help shed some light on the effect of previews on sales.

²<http://www.lessig.org/blog/archives/ag%2Cpdf.pdf>. While writing this, a settlement was reached between Google and the Author's Guild (<http://books.google.com/googlebooks/agreement/>). While that puts an end to that particular lawsuit it doesn't in any way diminish the importance of this topic.

2 Model

Every day thousands of people search for various information goods (music, books, movies etc.) through many different channels. Once they find a product that fits what they're looking for, they make a buy / no buy decision. Before their decision to buy they collect (or are "fed") information on the work in order to uncertainty about the quality of the good.³

A traditional view of pre-purchase information is that the potential buyer is uncertain about the true value of the good and is using pre-purchase information as a signal of the true value. In a simple scenario where the product can be of either high or low value, the model states that the buyer will accept some search costs in order to update their expected value of the product. If we flip this model around so that the firms offering the products now make a decision on how much pre-purchase information to provide we end up with the direct effect that the seller of a high value good (where the value is higher than the average value for all competing goods) will provide more information to convince the buyers of the value of the good (and motivate a higher price). The seller of a low value good can, in the short term, provide little or no information and expect buyers to make their purchase decision based on unconditional expected values (the mean value of all products). Signaling theory (e.g. [Spence, 1973](#); [Gibbons, 1992](#)) tells us that this is not an equilibrium solution since we know that the producers of high value goods will market their good extensively (and it is not in the interest of the seller of a low value good to spend money to convince the potential buyer that the product is of low value). Rather, the low value good will be priced lower to reflect the expected value conditional on only low value goods. Similarly, there is a sense that sometimes you could reduce your base of potential customers by providing more information about your product. This is analogous to the quality issue but can also be extended to different niches. The equilibrium is the same in that a) a rational (whoever rare) consumer would buy from you with the correct probability that your good is from the correct "niche" and you letting the consumers know ahead of time doesn't change the expected number of buyers, just shifts them around so that you end up with a better fit between customer and product. Because of the existence of reliable signals (our pre-purchase information) we avoid Akerman's "lemons solution" ([Akerlof, 1970](#)). In the end this model does very little to help with the analysis of information goods where a) the prices tend to be quite similar in the case of music and vary by factors other than quality in the case of books and, b) there is an

³It is true that marketing sometimes seems to be intended to inflate the perception of the real value, but we will ignore deliberately deceptive marketing here.

abundance of recommender information and feedback available that will prevent systematic cheating by the sellers. We should not expect the price to reflect the average expected value of the good in such a way that a book that is objectively better than a competing book is priced higher so as to make both reasonable choices.

In this paper I assume that the pre-purchase information is an unbiased predictor of the value of the work so that increasing the amount of this information improves the prediction of the value of the work (by reducing uncertainty about the true value) but does not, on average, change the expected value. Thus, I take a slightly different approach than the traditional one and adopt a concept introduced by [Bauer \(1960\)](#), perceived risk. Bauer introduces big negative costs (quite possibly non-linear) from making an erroneous “buy” decision. This leads to the conclusion that reducing uncertainty about the value of the product increases sales. The idea is quite simple. When we buy, we take the risk of suffering some kind of loss. Typically these losses are thought to come in a combination of four forms: financial, physical, time and psycho-social ([Mitchell, 1999](#)). Mitchell also claims that “...when risk is below a risk threshold, perceived risk theory has little explicatory power...”. He finds this to be especially true when (because) consumer involvement is low in the selection of the product. While books certainly are both low cost and, most likely, low risk, they are not a typical repeat purchase with low involvement.

I argue that previews are (essentially equivalent to) free samples when talking about information goods. [Roselius \(1971\)](#) examines the value of different means of reducing perceived risk. He finds that free samples is one of the more effective means. [Bawa and Shoemaker \(2004\)](#) builds a model for free samples that focus on the effect on the brand (repeat purchases) but also introduce the risk of the sample cannibalizing paid trial purchases.

A simple statistical incorporation of perceived risk into the model is to let expected value be just that, a mean of the perceived value. The perceived risk takes two forms if we adopt the view of [Kogan and Wallach \(1964\)](#) that risk is the combination of uncertainty and consequence of a negative outcome. The uncertainty component of the outcome is reflected in the variance of the expected value. The severity or danger of a negative outcome is captured by how certain we want to be that the true value is, in fact, larger than the price. This is analogous to standard hypothesis testing and the severity of buying the “wrong” good determines the rejection level for the test.

Just like [Bawa and Shoemaker \(2004\)](#) I recognize that previews can have a cannibalizing effect as well as a promotional effect since the previews transfer some of the value of the

work to the free, pre-purchase stage. Specifically, since I define previews to be a subset of the full product, there is always a chance that someone found the piece that he or she was looking for in the preview and no longer feels the need to buy the full product (even if they would have bought the product in the absence of the preview). In that scenario the preview is a perfect substitute for the full good when the price of the good is taken into account, i.e. the value of the preview is at least as large as the expected value of the good minus the price. For individual users these two effects are obviously binary (either you buy or you don't), but in aggregate you are bound to get two smooth curves with a total effect similar to the curve illustrated in Figure 1. A more rigorous mathematical approach to the model can be found in Appendix A.

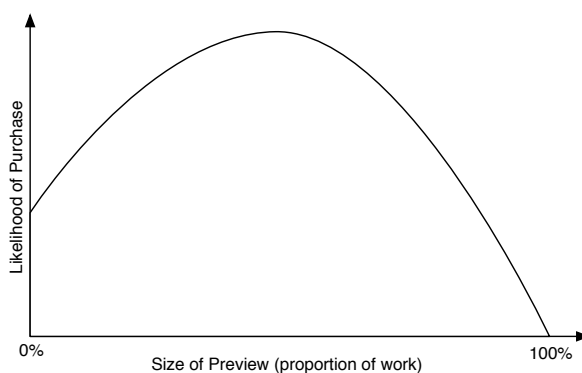


Figure 1: Demand as a function of previews

While the model doesn't really tell us how we should expect sales to be affected by previews, it seems clear that the sellers that use or contemplate these previews hope and expect the effect to be positive. In the specific case considered in this study, the previews are limited to a very small number of pages and, while there certainly is plenty of meta-information about the book in the form of reviews etc, we seem to be very far away from a case where 100% of the book is available for consumption prior to purchase. Thus we have strong reason to believe that we are left of the maximum in Figure 1 and we would expect sales to increase when *Search Inside!* is enabled. This leads us to the main hypothesis.

Hypothesis 1 *An increase in previews will increase sales as long as the total proportion of the good available for previewing is relatively small.*

The propensity to cannibalization will depend, to a large extent, on the type of content we are dealing with or, if you will, the genre. If we compare two books, one being a novel

and the other being an encyclopedia, it is clear that just seeing a few pages relevant to your search query will have a very different value. In the case of the novel, showing a few pages is unlikely to give the user the feel that she or he read the novel. While you might be able to get full satisfaction from a novel even though you skimmed over some paragraphs here and there, you would not appreciate a couple of missing pages. In the case of the encyclopedia, your demand might be 100% satisfied by a one page preview if it shows the paragraph with the answer you were looking for. In this context, encyclopedias are part of a larger set of reference books that are used to look up specific information. In between novels and reference literature we can, for example, expect to find a group of “text books” where the relevant (to the individual user) section of text is larger than in the reference literature but still much smaller than the “all or nothing” that could be expected for novels. It can be hypothesized that the genre of a book would have a direct impact on the likelihood of cannibalization. In the context of this experiment the limitations of the data force me to reduce this to a more narrow hypothesis:

Hypothesis 2 *Previews are more likely to cannibalize sales for reference literature than any other genre of books.*

It is important to recognize that the model is focusing on previews. Specifically it is dealing with the conversion ratio of “looks” or “visits” to sales. After all, the preview has no impact on the consumer becoming aware of the good but only takes effect once the consumer is trying to evaluate whether or not to buy this particular good.

3 Experimental Design

For the experiment, I was initially cooperating with O’Reilly Media (ORA). Together we compiled a list of 400 books that they were selling through a variety of sources, although predominantly through Amazon.com. While a few of these books had *Look Inside!* enabled, none of them provided Amazon’s *Search Inside!*. The experiment therefore allows us to observe the effects of enabling *Search Inside!*.⁴ It is important to know that we are not going

⁴*Look Inside!* allows customers to browse sample pages and do additional searches inside a particular book. *Search Inside!* extends this and also allows users searching on Amazon.com to find matches based on text inside the book as opposed to only using meta data. From <http://www.amazon.com/gp/help/customer/display.html?ie=UTF8&nodeId=13685651&qid=1228598632&sr=2-1>. *Look Inside!* and *Search Inside!* are registered trademarks of Amazon.com, Inc.

from a state of no pre-purchase information to the treatment state. Library records are available for all books (these include title, author etc.). Amazon also provides a wealth of information about all books they sell whether or not they have *Search Inside!* enabled. This includes a picture of the front page, an abstract, customer reviews when available, Amazon sales rank etc. This experiment only measures the marginal effect of adding *Search Inside!*.

As was discussed on page 7, the relevant unit of measurement for the model I want to test is previews or page views (for each book). Neither of these are available to me and I will have to use sales/book as a proxy for sales/visit. Further, the sales data used is not pure Amazon sales although Amazon contributes the vast majority of the sales. I explicitly assume that Sales/Book is an unbiased estimator of Sales/Visit to the books Amazon page.

The more important problem is that we only expect previews to impact the conversion of visits to sales. However, *Search Inside!* should improve the ability of a potential customer to find the book's page in the first place, thereby increasing visits. Unlike the conversion case where the model calls for two opposed effects, marketing and cannibalization, there is no theoretical negative effect in the discovery phase. This does not impact our hypotheses if we assume that the impact of *Search Inside!* on visits/views is equal in expectation within all pairs of books. The most likely case where this could be problematic is when testing Hypothesis 2 since we now have to assume that reference literature has the same impact from the treatment as all other books. If Hypothesis 2 is true (and these books are more prone to cannibalization) it may also be true that more people will be attracted by the *Search Inside!* feature when looking at comparable books, thus increasing the number of visits relative to all other books. Unfortunately we cannot control for it *a priori* nor test for it *a posteriori*.

The data is composed of a panel with 400 books on one axis and weekly sales data on the other. This allows the use of a variety of estimation methods for the treatment effect:

- Before vs. After (During)
- Treatment vs. Control
- Difference-in-Difference

Difference-in-Difference (DiD) is the “catch all” method since it allows us to correct for any existing differences between the treatment and control groups as well as temporal effects, at the cost of assuming additive separability in time. Specifically the former means that we

can correct for the difference between treatment (T) and control (C) when $E[T - C] \neq 0$ and the latter allows us to accommodate changes in the sales from the before to the treatment period that were unrelated to the treatment itself (external shocks, seasonal effects etc.). That said, it is also less efficient than the other two options since it introduces larger variance. The leading candidate in terms of simplicity is always Before/After since you don't need a control group. In terms of efficiency the gold standard is the Treatment/Control setup.

I ran the experiment in such a way that I could collect complete information for a DiD model, but created our treatment and control groups in such a way that I can get results by simply looking at the treatment and control data from the treatment period. Simple randomization over the 400 books does (essentially) achieve the $E[T - C] = 0$ criteria. However, there are a number of strategies to reduce the variance of the estimate and thereby increase the power of the tests. Finally, while we can control the start of the treatment, we cannot remove it (*Search Inside!* cannot easily be disabled) which leads to the deployment strategy shown in Table 1 where treatment is first applied to one group (in period one) and then also applied to the second group in period two.

Table 1: Treatment Application

Period	0	1	2
Group A	No	Yes	Yes
Group B	No	No	Yes

The rest of this section is focused on the setup of the experiment based on the historical sales data available for our dataset.

3.1 Treatment/Control Selection

For the 400 books in question I have weekly sales data going back to September 2006 and up to January 2008 with average weekly sales varying from 0.14 to 450.⁵ Most of the books were published before the starting date of the data series but 31 were added later (the panels are not balanced). That said, the latest addition was published in February 2007, so there is close to one year or more of historical data for every book. The data is also somewhat sparse in that most of these books have relatively small sales. With a large number of the books

⁵Mean: 35.51, Standard Deviation: 55.09

selling less than one book per week on average, there are many weeks where some books had 0 sales. This becomes problematic when working with weekly data and a log model.

Sales distributions vary in a number of ways...

- Over time (they tend to sell more soon after they are released with sales falling off over time).
- Within books, common count data results in a skewness to the right with the mean sales higher than the median.
- Across seasons.
- Across books. A few books sell a lot which makes the mean higher than the median.

[Sorensen \(2007\)](#) finds that new books sell more. The same feeling was corroborated by *O'Reilly* and is illustrated in our data in [Figure 2](#). The bars represent the proportion of the books in the sample of a certain age (on the right axis) and the line shows the average sales (left axis). The spike in sales for the really old books most likely comes from new editions being released. While I considered incorporating age into the pairing algorithm it is also worth noting that, by the time we expected to run the experiment (February 2008), no book would be younger than one year and the major sales spike resulting from a new release would therefore have passed.⁶

With sales being “count data” one would expect the weekly sales to be skewed and follow something reminiscent of a Poisson distribution within each book. [Figure 3](#) shows the weekly sales for each book normalized by the average sales for that book. If each book followed a Poisson distribution, this aggregation of normalized sales should also show that. This is in fact what we see in [Figure 3](#), but with a spike at 0. In [Figures 4](#) and [5](#) I divide the dataset into books with average weekly sales larger than five and smaller (or equal) to five. For the larger than five set we do indeed see something closely resembling a poisson distribution whereas the lower than five series contains 90% zeros. Rather than worry about the distribution of the sales for the individual books, be that Poisson or, more likely Negative Binomial, I create pairs of books and assign a treatment and a control within each pair, thereby getting a symmetric distribution.

[Figure 6](#) shows that sales do have a seasonality component, although the types of books ORA sells (professional literature) most likely have far less seasonality than, for example,

⁶The experiment was further delayed and didn't actually start until May.

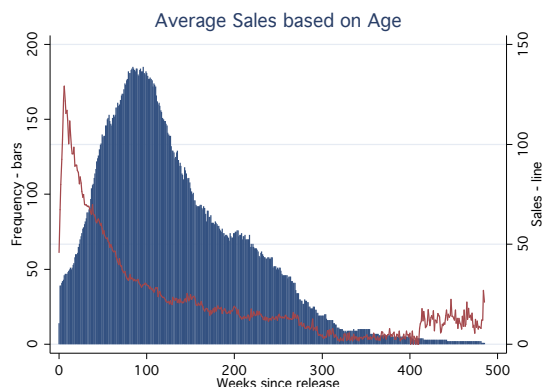


Figure 2: Do books sell more when they are new?

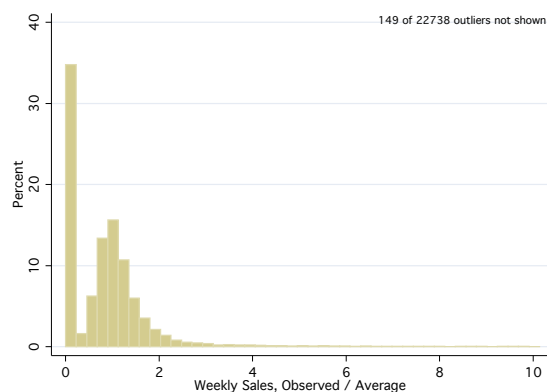


Figure 3: Weekly Sales by Book, normalized by mean sales per book

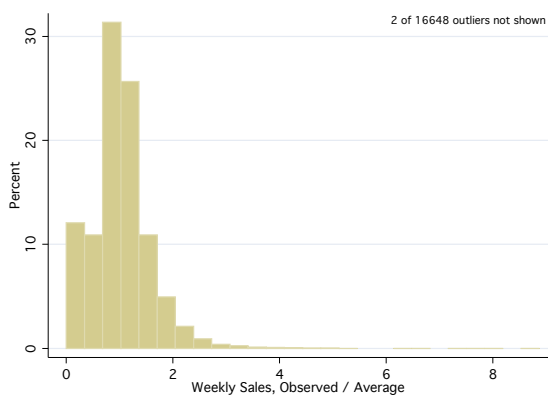


Figure 4: Weekly Sales by Book if mean sales > 5

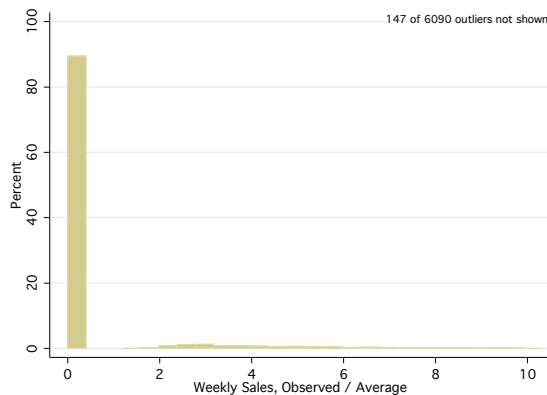


Figure 5: Weekly Sales by Book if mean sales ≤ 5

novels and text books. Still there is a dip during the summer as well as spikes for Thanksgiving and Christmas. This is of very little importance in any Treatment/Control or Difference-in-Difference estimation.

Finally there is concern about the skewed distribution of sales between books. While it is not a problem getting an expected difference between Treatment and Control to zero, the books with higher sales obviously have much larger variance than the low sellers as can be clearly seen in Figure 7. It is natural to expect the effect to be multiplicative, leading us to use a log-linear model for sales. This is shown in Figure 8. The major downside is a loss of observations since many books have weeks with 0 sales. Figure 8 conveniently hides this by

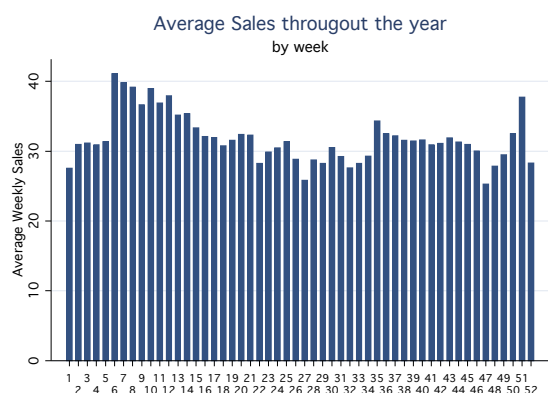


Figure 6: There are seasonal effects.

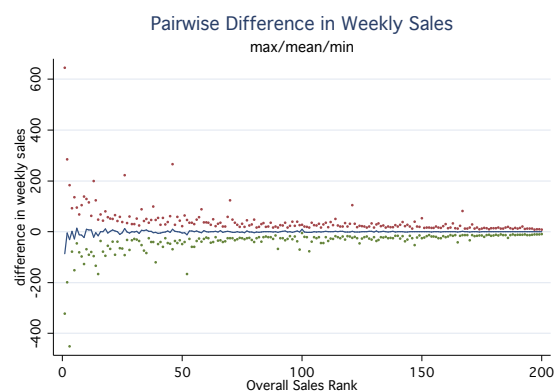


Figure 7: Variance is relative to sales.

just omitting those weeks (rather than throwing out the whole pair), but that ties into one of the obvious solutions to this problem. By aggregating sales for a longer period (be that before or during) we reduce the risk of seeing any zeros.⁷

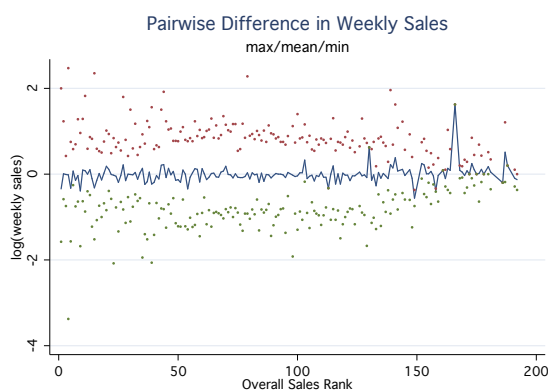


Figure 8: Log-linear model reduces heteroskedasticity

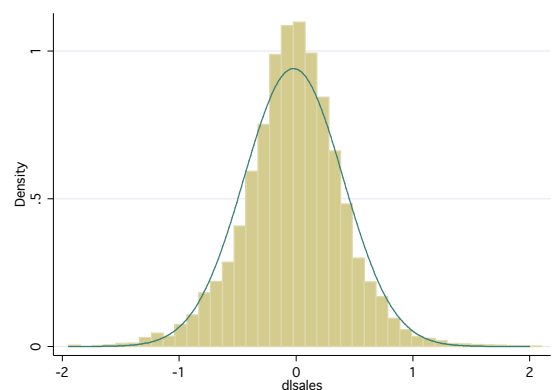


Figure 9: Distribution of $\log(T) - \log(C)$ compared to a normal distribution.

Thus a naïve test of means can be improved upon by using log of sales to reduce heteroskedasticity and use paired sales to create an easy to analyze, symmetric distribution. Figure 9 shows how well the difference of logs model fits the normal distribution.

Recognizing that the goal is to make the difference of $T - C$ as efficient as possible I want

⁷As you will see in section 4, in the end I only had to throw out two pairs for this specific reason.

to make the matching/ranking as good as possible. I started with a very simple ranking based on 95% trimmed mean weekly sales while separating books with one or more 0 sale weeks from those without (to avoid “tainting” of the pairs). The goal is, obviously, to find an estimator that would be a good predictor for the sales during the treatment period. I then went on to use more elaborate models that incorporated the age of each book as well as seasonality. In the end these more complex models didn’t change the ranking, and thereby the pairing, by much. Instead most of the age effect can be removed from the pairing by only looking at the most recent weeks (and we have already determined that there is no need to worry about seasonality when running treatment/control and diff-in-diff models). I ended up with a very simple ranking based on the mean sales over the last 12 weeks of our historical sales data set.

3.2 Power Calculation

The next step is to try to determine how long to run the experiment (the other variable, number of observations, is not available since we are already using the 400 top sellers and have a significant 0 problem even in among those). Since I have access to historical sales data I can, rather easily, simulate experiments of different lengths and use that to determine the power.

Specifically, I generated a distribution of confidence intervals for the null hypothesis (effect of treatment = 0) and then calculated what proportion of that distribution would be significant should there exist a real effect of the treatment of X%. The following assumptions were used:

1. Create pairs using mean weekly sales and assign treatment and control at random within each pair.
2. Pick w weeks (4, 6 and 8) as the treatment period without replacement within each pair. For each pair, calculate the difference of log mean sales,
$$Diff_i = \log\left(\frac{1}{w} \sum_{i=0}^w (sales_{i,T})\right) - \log\left(\frac{1}{w} \sum_{i=0}^w (sales_{i,C})\right).$$
3. Run a two-sided t-test of $Diff_i = 0$ and calculate the lower bound confidence interval at 95%.
4. Run 2000 simulations of this and use the distribution of confidence intervals to calculate the power at different levels of actual treatment effects. I.e. in what portion of the

simulations would a p% difference (increase) in means be significant?

The result for a simulated treatment/control experiment as outlined above and treatment periods of 4, 6 and 8 weeks can be found in Figure 10 and Table 2. Table 2 shows the breakdown of the effect between standard deviation within each pair and number of observations. A longer experiment period leads to fewer dropped pairs due to 0 sales (for which there is no logarithm) and therefore more observations. This is outweighed by a larger observed variance within pairs as the experiment period gets longer, leading to a net loss in power from longer treatment periods.⁸ While we could, most likely, decrease the variance and thereby improve the power of the experiment by including more books, the sample size was restricted by the collection of suitable titles from the publisher and that avenue was therefore closed to us.

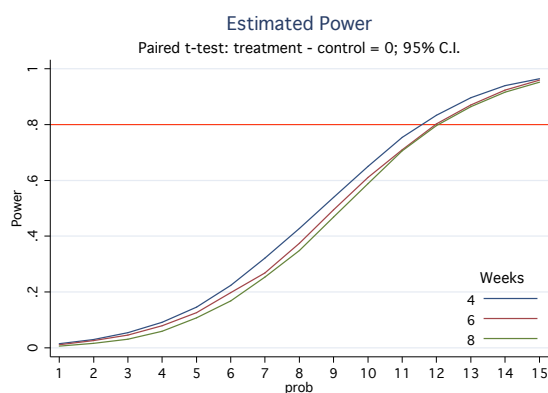


Figure 10: Power Simulation

Table 2: Difference in Log Sales over 2000 simulations.

	4 weeks	6 weeks	8 weeks
#Pairs	145.4	152.5	157.7
Mean	-0.001	-0.001	0.001
Standard Error (of mean)	0.051	0.054	0.056
Standard Deviation (within pairs)	0.617	0.663	0.705

⁸This is quite unintuitive and I have no explanation for this. The effect is, however, quite small.

4 Data

In preparation for the experiment I started tracking sales of 400 titles. These titles do not represent a random sample, but were the top 400 sellers from the publisher at the time. The cutoff at 400 was chosen based on the average sale of the marginal book. We started with 200 titles but decided to make try to use a large sample. At 400, the marginal books have such low sales that it becomes hard to extract any information since we can only observe changes of minimum one book (and with low sales that becomes a very large percentage change in sales). Further, you cannot observe a decrease in sales if the sales are already zero. I also elected not to include 20 books available as ebooks on Amazon.com since I was concerned that they would behave too differently and there would be too few observations to really estimate an effect of ebooks specifically. Thus the 400 books is, for all intents and purposes, the population of relevant books from the publisher.

All books from the publisher are of a technical (IT) nature but range from pure reference in the *Cookbook*, *Hacks* and *Pocket Reference* series to the more textbook like consumer series with books such as the *Missing Manual* series.

The determination of whether or not a book had received the treatment was made by running a daily perl script that would download the amazon.com page for each book in the sample and then parse the HTML to check for a link to *Search Inside!*. While the scraping is accurate there is some concern that different users of Amazon.com might see different pages. I made a point of randomly checking books each day from a completely different computer (different IP address and without cookies enabled). As can be seen in the reasons for eliminating pairs below, two books did have a strange behaviour in this regard but no other anomalies were uncovered.

The sales data was acquired from Nielsen Bookscan and covers sales through all sources.⁹ Since, as was mentioned earlier, most of the books from the publisher are sold through Amazon.com, I assume that it is a good proxy for the sales through Amazon.com even though it does mean that any effect will be slightly underestimated. The actual start and end date of the experiment was not known to me, thus I set up a script to monitor Amazon.com daily and collect data on when *Search Inside!* was enabled for each of the ISBNs in the study.

The experiment started in week 12 and ended week 20, 2008. Since I only have weekly sales data (rather than daily) I throw out weeks with mixed states (first and last week) to

⁹Nielsen BookScan provides weekly point-of-sale data for books sold in the United States and function as a central clearinghouse for the US book industry.

ensure that the difference between treatment and control isn't "contaminated".

Pairs of books are eliminated for the following reasons.

- 1 book had no sales data.
- 10 pairs had no difference in treatment/control during the treatment period (either both or neither book had *Search Inside!* enabled).
- 2 books showed up as having *Search Inside!* enabled even before the experiment started. While this seemed to depend on who was looking at the Amazon site, it made it impossible to verify when *Search Inside!* was actually enabled for them.
- The treatment books within 2 pairs had inconsistent treatment during the period (the Amazon scan shows them without *Search Inside!* for a few days of the period). I have tested both with and without these pairs and don't find much effects from keeping them. They are thus eliminated from the final analysis.
- Finally, 2 pairs had at least one book with no sales during the experiment period resulting in invalid log of sales. Rather than using an adjustment (adding 1 to all sales for example), I eliminate those pairs.

After this cleaning we are left with 183 pairs out of the original 200 pairs. Of these, not all have data for the full treatment period (8 books had the treatment applied late). Since the books with the late treatment had higher than average mean sales, I report estimations both with and without them. Having fewer weeks to use for the estimator only affects the precision of the average sales (variance), not the mean value since the mean is an unbiased estimator whether it is calculated over 1 week or 7 weeks. Figure 11 shows the distribution of the within-pair difference between logged weekly sales for the treatment period (after the cleaning outlined above).

5 Analysis

Following the logic described in the model, we can test hypothesis one with a simple paired t-test. This is, of course, equivalent to an OLS regression with pair dummies, but the regression allows us to add other covariates to the model later. I therefore use the OLS setup:

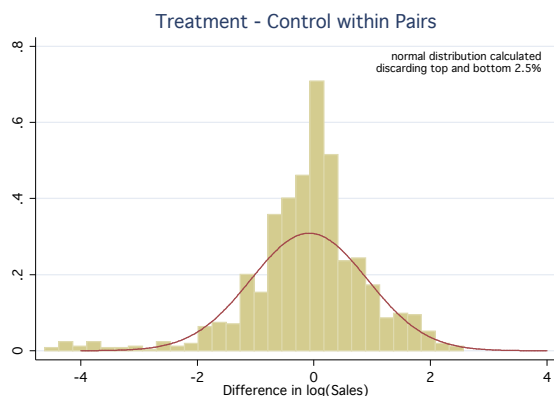


Figure 11: Distribution of $\log(T) - \log(C)$ compared to a normal distribution.

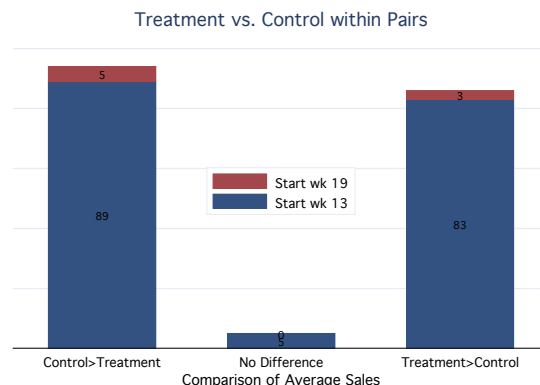


Figure 12: Sign of $\log(T) - \log(C)$ (non-parametric).

$$\log(\bar{S}_i) = \alpha + \beta_1 TD_i + \sum_{k=1}^{200} \beta_k PD_{k,i} + \epsilon_i$$

\bar{S}_i is the average weekly sales during the treatment period for book i , k is the pair $\{1, 2, \dots, 200\}$ and TD_i and PD_k are dummy variables for Treatment and Pair ID respectively. If book 1 and 2 are both part of pair 4 and book 1 receives the treatment, $TD_1 = 1$ and $PD_{4,1}$ and $PD_{4,2} = 1$, all other $PD_{4,k} = 0$. The coefficient on the TD is the multiplicative change in sales observed from the treatment and the null hypothesis that $\beta_1 = 0$.

The regression coefficients are the combined effect of the two contradictory effects outlined in the model and reflects the marginal effect of the treatment. As discussed earlier, it is impossible to decide how much value is provided by the previews as a proportion of the total value of the work (unless the preview includes full access to the work), thus we don't know where on the curve we are a priori. If we accept the shape of the curve in Figure 1, we can assume that a positive impact (coefficient) means we are left of the optimum and a negative coefficient places is right of the same.

Table 3 shows the results from three separate OLS regressions. Models 1 and 2 use pair dummies and the difference is that model 2 excludes the “late entry” treatment pairs identified earlier. It is clear that omitting them does little to the results (although the coefficient is smaller). The unpaired test (model 3) is included for reference to verify that using the paired setup didn't create any large anomalies.

What little effect there is is negative, i.e. *Search Inside!* reduces the expected sales. Since it is non-significant we can either decide that there isn't enough data to say where on the curve we are or, since the curve has a maximum, determine that we are near that maximum both with and without the treatment.

Table 3: Treatment effect from treatment/control estimation.

	(1)	(2)	(3)
Effect from Treatment	-0.043 (0.069)	-0.020 (0.068)	-0.043 (0.142)
Pair Dummies	Yes	Yes	No
Include "Late" Treatment	Yes	No	Yes
Observations	366	350	366
Adjusted R^2	0.76	0.78	-0.00
Significance levels:	† : 10%	* : 5%	** : 1%
Standard Errors in parenthesis			

As was discussed in section 3, the treatment/control setup depends on the assumption that the difference between the group would, on average, be zero absent any treatment. Our setup should ensure that. However, since we had to eliminate some pairs for various reasons (that may or may not be random) it is reasonable to test this assumption. We have two periods where there is no difference in treatment between the two groups, 0 and 2. Period 0 is pure and unadulterated. In period 2 we have one group that have already had *Search Inside!* enabled for one period whereas the other group is "freshly enabled". If there is any delayed effect from enabling *SI*, the assumption of a zero difference doesn't hold. Table 4 shows the results for periods 0 and 2 by themselves and both tested jointly (with interaction effects to control for any lag effect). The difference is significant at 10% and, while that may not be enough to reject the underlying assumption, it does raise a warning flag.

5.1 Difference-in-Difference Model

With perfect randomization there is no reason to test anything but treatment vs. control. However, since we could not use the full 200 pair sample that was randomized, we may not have a perfectly randomized sample. As was discussed in the previous section, there are

Table 4: Difference between groups A and B.

	Period 0 b/se	Period 2 b/se	Periods 0 and 2 b/se
Group A	-0.103 [†] (0.057)	-0.068 (0.072)	-0.098 [†] (0.055)
P2			-0.346 ^{**} (0.052)
P2 × Group A			0.020 (0.077)
Observations	393	390	783
Adjusted R^2	0.85	0.78	0.86

Significance levels: † : 10% * : 5% ** : 1%
Standard Errors in parenthesis

indications that there could be a systematic difference between the groups A and B (the treatment and control groups). Using a difference-in-difference model allows me to control for systematic (pre-existing) differences between the treatment and control groups. The main assumptions are that any underlying difference between the groups is constant over all periods and that any period difference affects both groups the same.

I run a simple OLS regression with pair, treatment and a full set of panel dummies (group, period and interactions between the two). I test it three ways, period 0 vs. 1, 1 vs. 2 and all three periods. The treatment of the two groups by period is shown in Table 5 where 1 indicates *Search Inside!* enabled. During period 0 neither group is treated, during period 1 *Search Inside!* is enabled for Group A and during period 2 it is also enabled for the Group B group. Since I didn't see any significant effect from the inclusion of the "late treatment" pairs in the basic model (see Table 3) I am excluding those pairs from this analysis.¹⁰

The results are reported in Table 6. Since the observations within each pair are, most certainly, dependent we use the Huber/White estimator of variance clustered on pairs (Greene, 2000; StataCorp, 2001). Model (1) only estimates the effect of *Search Inside!* on Group A, (2) only estimates the effect on Group B and model (3) has the combined effect. The treatment effect is now positive but insignificant. The negative effect observed in the treat-

¹⁰While it would be easy to include them, it muddles the definition of the periods since different pairs would have different start and end dates for each period.

Table 5: Difference-in-Difference setup.

	Period 0	Period 1	Period 2
Group A	0	1	1
Group B	0	0	1

ment/control model is absorbed by the dummy for Group A that show significantly lower sales in both models one and three. The model also shows that sales decline over time which is consistent with our assumption that older books sell less but should not otherwise impact the experiment. In period 2, Group A shows an (insignificant) increase in sales over Group B over and above the baseline difference identified by the Group A dummy. If this difference is real it could mean that there is a lagged positive effect from treatment. That is, enabling *Search Inside!* affects sales more as time goes by and the full effect isn't realized immediately. This seems reasonable.

While the sign for the treatment effect changed in the D-in-D model compared to the T/C model, neither effects are significant and the null in Hypothesis 1, that previews have no impact on sales, cannot be rejected.

5.2 Are reference books different?

Next I look at the possibility that reference books could be different. The idea is that there are certain books that are used to look up a very specific answer to a specific question and that such an answer could easily be displayed in full even when only a limited number of pages are returned after a search (from services such as *Amazon's Search Inside!* or *Google Book Search*). Together with *O'Reilly* I identify two sub-categories of books in the sample that should be considered reference books and run the paired regression with full set of interaction dummies. Table 8 includes both a Treatment/Control model and a three period Difference-in-Difference model (same setup as Table 6, model 3). Robust standard errors are used in both models and the DiD model also clusters the standard errors on pairs.

The 72 reference books are distributed fairly evenly between treatment and control as shown in Table 7. The results reported in Table 8 show that including controls for reference literature does not change the direction of the effects found in the baseline models. It does however strengthen those results, especially in the T/C case where treatment now shows a

Table 6: Treatment Effect from Difference-in-Difference estimation.

	(1) 0 vs 1	(2) 1 vs 2	(3) 0 vs 1 vs 2
Treatment	0.085 (0.054)	0.035 (0.054)	0.079 (0.052)
Group A	-0.112 [†] (0.065)	-0.086 (0.083)	-0.125* (0.060)
Period 1	-0.260** (0.035)		-0.254** (0.033)
Period 2		-0.145** (0.040)	-0.445** (0.078)
A × P2			0.047 (0.063)
N (Pairs * Periods)	368 * 2	360 * 2	358 * 3
Adjusted R^2	0.85	0.82	0.85

Significance levels: † : 10% * : 5% ** : 1%

Robust Standard Errors in parenthesis clustered by pair

Table 7: Distribution of reference and other books between groups A and B.

	Regular	Reference	Total
Group A	145	38	183
Group B	149	34	183
Total	294	72	366

significant and very big (-17%) effect. As before, the D-in-D model attributes this effect to a baseline difference between groups A and B.

The variable “Reference” shows that reference literature sell less than other books. This is not really that interesting since we’re making no attempt to explain which books sell more than others outside of the use of *Search Inside!*. SI’s impact on reference literature is shown by the interaction effect of treatment and reference. In both models this shows that reference literature gets a larger boost from *Search Inside!* than the average book in the sample. This comes as quite a surprise since, if any type of book should suffer from the replacement effect

Table 8: Do reference books behave differently?

	T/C b/se	D-in-D b/se
Treatment	-0.170 [†] (0.088)	0.054 (0.055)
Reference	-0.368* (0.149)	-0.162 [†] (0.092)
Reference × Treatment	0.672** (0.231)	0.125 [†] (0.071)
Group A		-0.123* (0.060)
Period 1		-0.254** (0.033)
Period 2		-0.444** (0.078)
Group A × P2		0.046 (0.063)
Observations	368	1074
Adjusted R^2	0.77	0.85

Significance levels: † : 10% * : 5% ** : 1%

Robust Standard Errors in parenthesis

discussed in the model section, it would be books of the reference kind. However, as was discussed on page 8, there is reason to believe that reference literature could benefit more from improved *search results* and therefore get more visits. This in turn could lead to higher sales even if the conversion rate benefit from the treatment is smaller than for other genres. Since the total effect on sales is the product of the effect per preview and the number of previews, you could end up with the result we see (that reference literature benefit more from previews) while still being consistent with Hypothesis 2.

That said, I have found no evidence that allows me to reject Hypothesis 2.

6 Discussion

The insignificant effects found in the experiment are somewhat unexpected and leave plenty of room for speculation. A number of possible reasons (and criticisms of the experiment) comes to mind. First of all, it is possible that the *Search Inside!* feature simply doesn't provide enough information to have a significant effect on the sales decision (in terms of lowering uncertainty). Second, it is possible that we are somewhere around the maximum in the model where the marketing value of the *Search Inside!* is offset by the previews cannibalizing on sales. Finally, it may simply be that the sample is too small (the variance too great) to be able to expect significant results.

The limitations of the sample (size and variance) is the most likely explanation for the outcome and is collaborated both by our pre-experiment power analysis discussed in section 3.2 and also by the post-experiment power analysis reported below. Given the results reported in Table 8 where reference literature was found to get more benefit from *Search Inside!* than other books, it seems likely that the lack of significant effect is also affected by the actual value contributed by *Search Inside!* being low for most buyers.

It is also important to recognize that the sample, while not all that small per se, is from one fairly narrow category of books all offered by the same publisher. It is very likely that the results are driven, one way or another, by that fact and it is important not to make too far reaching conclusions based on one experiment.

In section 3 I noted that the power limited by the number of books that could be included in the experiment. With estimated standard deviations for the difference within pairs, I can estimate power if I had access to a larger sample. Figure 13 shows the power curve based on a two-sided 95% confidence interval for the experiment and an assumed normal distribution for the population with a standard deviation per pair equal to the sample standard error in the experiment. The figure shows that we would need to more than double our sample size to get power above 80% in case of a 10% effect and that it is all but impossible to get the power that high for our observed 4% effect (from model 1, Table 3) and even an 8% effect (if the true effect is more like model 3, Table 6) would require an extremely large sample.

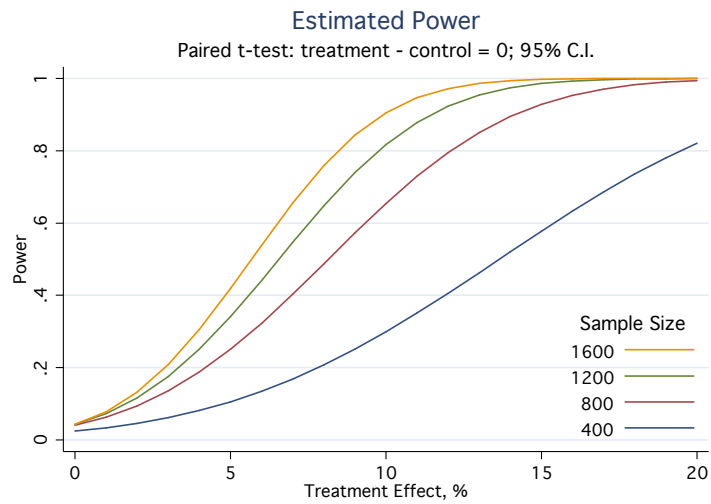


Figure 13: Power calculation for various sample sizes.

A Formalizing the Previews model

A.1 Buying under Perceived Risk

The individual user j will make the purchase of good i if his expected value exceeds the cost of the good (P_i) with probability $p_{i,j}$ based on uncertainty the expected value measured by variance.¹¹ we are going to assume, for the sake of modelling, that the true value ($V_{i,j}$) is distributed around the expected value ($E[V_{i,j}]$) with variance ($v_{i,j}$) and following some distribution function $f(E[V_{i,j}])$. Thus I assume that the user i will purchase good j when:

$$\frac{E[V_{i,j}] - P}{v_{i,j}} \geq CDF_f(p_{i,j}) \quad (1)$$

As was discussed before, the value of v depends on the uncertainty about the true value of the good. In this model, this uncertainty is inversely proportional to the amount of pre-purchase information available (pp_i), i.e. previews.

To get a bit more structure we first define pp_i as the proportion of good i available as the preview preview. Let $V_{i,j}^{pp}(pp_i)$ be the value to consumer j of the preview pp_i . We now define pp_i as the proportion of the average value of the preview for all consumers (J), V_i^{pp} where

$$V_i^{pp} = \frac{1}{J} \sum_j V_{i,j}^{pp}(pp_i) \quad (2)$$

and

$$pp_i = \frac{V_{i,j}^{pp}}{\frac{1}{J} \sum_j V_{i,j}} \quad (3)$$

This allows us to order all possible previews for good i so that $\frac{\partial V_i^{pp}}{\partial pp_i} > 0$ and $V_i^{pp} \in [0, 1]$.¹²

The aggregate effect of the previews on sales is determined by the aggregate probability that inequality 1 holds given a specific “amount” of previews, or similarly, the proportion of all users for which inequality 1 holds at a given level of previews. We have already assumed V_i^{pp} to be increasing in pp_i and we are implicitly assuming that perceived risk, as captured by v is decreasing in V_i^{pp} (in aggregate), i.e. $\frac{\partial v_i}{\partial V_i^{pp}} \leq 0$. This is a direct effect of our assumption

¹¹Note that I assume p to be at least 50%. I.e. you want a better than 50/50 chance of the value being higher than the price.

¹²The value of the previews will certainly not be a monotone increasing function for individual consumers, but we are interested in the aggregate effect.

that more information cannot, on average, increase uncertainty. This leaves us with an upward sloping demand function in pp_i .

A.2 Sample cannibalizing sales

Cannibalization occurs if the potential buyer gets enough value from the preview to offset the expected surplus from buying the good.

$$E[V_{i,j}] - P_j \leq V_{i,j}^{pp}(pp_i) \quad (4)$$

Again we ask ourselves how the aggregate demand for the good depends on pp_i . In equation 4 only the right hand side depends on the amount of previews and we already know that $\frac{\partial V_i^{pp}}{\partial pp_i} > 0$ in aggregation. Thus the aggregate probability that equation 4 holds is strictly decreasing in pp_i .

A.3 Combined model

At the level of the individual consumer, we are left with three possible outcomes. First of all the consumer can decide to buy the good, second he or she can decide that the preview satisfies his/her demand and, lastly, the consumer can decide to hold off on the purchase and wait or search for better information. Inequalities 5 and 6 shows the combined conditions for purchase. Note that inequality 6 requires the left hand side to be strictly less than the right hand side. That is because the right hand side precedes the left hand side in time (the customer “consumes” the preview before deciding whether or not to buy).

$$\frac{E[V_{i,j}] - P_j}{v(pp_i)} \geq CDF_f(p_{i,j}) \quad (5)$$

$$E[V_{i,j}] - P_j > V_{i,j}^{pp}(pp_i) \quad (6)$$

In aggregation we have two functions of pp_i . p^m is the proportion of consumers buyers for which there is enough pre-purchase information to make them feel confident in buying, i.e. sufficient marketing. p^c , the cannibalization function, is the proportion of consumer for which the available pre-purchase information will cannibalize potential sales.

$$p^m(pp_i) = p \left(\frac{E[V_{i,j}] - P_j}{v(pp_i)} - CDF_f(p_{i,j}) \geq 0 \mid pp_i \right) \quad (7)$$

$$p^c(pp_i) = p(E[V_{i,j}] - P_j - V_{i,j}^{pp}(pp_i) \leq 0 \mid pp_i) \quad (8)$$

While we do not know the exact shape of either distribution function we do know the slopes and, thus, the general shape of the combined demand function.

$$\frac{\partial p^m(pp_i)}{\partial pp_i} > 0 \quad (9)$$

$$\frac{\partial p^c(pp_i)}{\partial pp_i} > 0 \quad (10)$$

Since sales will occur only if both inequalities 5 and 6 hold, the combined demand function is simply the net effect of the two.

$$p^m(pp_i) - p^c(pp_i) \quad (11)$$

Since both components are increasing in pp_i we are likely to end up with a demand function similar to the one shown in figure 1 where sales are increasing in pp_i until the cannibalization becomes too great, after which further pp_i cause the sales to decrease. It is, of course, possible that $p^c(pp_i) > p^m(pp_i)$ for all pp_i . In that case no sales occur at any level of pre-purchase information (the curve is flat = 0). The opposite cannot occur in this model (positive sales regardless of pre-purchase information) since, if we give away 100% of the good, the consumer has all the value at no cost (which is a higher surplus than if he had to pay the purchase price).

References

- George A. Akerlof. The market for 'lemons': Quality uncertainty and the market mechanism. *Quarterly Journal of Economics*, 84(3):488–500, Aug 1970. URL <http://links.jstor.org/sici?sici=0033-5533%28197008%2984%3A3%3C488%3ATMF%22QU%3E2.0.CO%3B2-6>.
- Raymond A. Bauer. Consumer behavior as risk taking. In R. S. Hancock, editor, *Dynamic marketing for a changing world: proceedings of the 43rd national conference of the American Marketing Association*, pages 389–398. AMA, Chicago, June 1960.
- Kapil Bawa and Robert Shoemaker. The Effects of Free Sample Promotions on Incremental Brand Sales. *MARKETING SCIENCE*, 23(3):345–363, 2004. doi: 10.1287/mksc.1030.0052. URL <http://mktsci.journal.informs.org/cgi/content/abstract/23/3/345>.
- BBC News. Downloading 'myths' challenged. URL <http://newsvote.bbc.co.uk/mpapps/pagetools/print/news.bbc.co.uk/1/hi/technology/4718249.stm>.
- John Borland. Study: File sharing boosts music sales. URL http://news.com.com/Study+File+sharing+boosts+music+sales/2100-1023_3-898813.html.
- Donald F. Cox and Stuart U. Rich. Perceived risk and consumer decision-making: The case of telephone shopping. *Journal of Marketing Research*, 1(4):32–39, Nov. 1964. URL <http://links.jstor.org/sici?sici=0022-2437%28196411%291%3A4%3C32%3APRACDT%3E2.0.CO%3B2-X>.
- Robert Gibbons. *Game Theory for Applied Economists*. Princeton University Press, Princeton, New Jersey, 1992.
- William H. Greene. *Econometric Analysis, 4th Edition*. Prentice Hall, Upper Saddle River, New Jersey, 2000.
- Nathan Kogan and Michael A. Wallach. *Risk-taking: A Study in Cognition and Personality*. Holt, Rhinehart & Winston, New York, NY, 1964.
- Stan Liebowitz. Record Sales, MPR downloads, and the Annihilation Hypothesis. August 2002.
- Richard Menta. Did napster take radiohead's new album to number 1? URL <http://www.mp3newswire.net/stories/2000/radiohead.html>.
- Norbert J. Michel. Internet file sharing: The evidence so far and what it means for the future. Technical report, The Heritage Foundation, August 2004. URL <http://www.heritage.org/Research/InternetandTechnology/bg1790.cfm>.

- Vincent-Wayne Mitchell. Consumer perceived risk: conceptualisations and models. *European Journal of Marketing*, 33(1/2):163–195, 1999.
- Felix Oberholzer and Koleman Strumpf. The effect of file sharing on recorded music: An empirical analysis. March 2004. URL http://www.unc.edu/~cigar/papers/FileSharing_March2004.pdf.
- Ted Roselius. Consumer rankings of risk reduction methods. *Journal of Marketing*, 35(1):56–61, Jan 1971. URL [http://links.jstor.org/sici?sici=0022-2429\(197101\)35%3A1%3C56%3ACRORRM%3E2.O.CO%3B2-F](http://links.jstor.org/sici?sici=0022-2429(197101)35%3A1%3C56%3ACRORRM%3E2.O.CO%3B2-F).
- Michael D. Smith and Rahul Telang. Competing with free: The impact of movie broadcasts on dvd sales and internet piracy. July 2008. URL <http://ssrn.com/abstract=1028306>.
- Alan T. Sorensen. Bestseller lists and product variety. *The journal of Industrial Economics*, LV(4):715–728, December 2007.
- A. Michael Spence. Job market signaling. *The Quarterly Journal of Economics*, 87(3):355–374, August 1973. URL <http://ideas.repec.org/a/tpr/qjecon/v87y1973i3p355-74.html>.
- StataCorp. *Stata User's Guide, Release 7*. Stata Corporation, College Station, Texas, 7 edition, 2001.