THE IMPACT OF THE INTERNET ON THE SALES DISTRIBUTION: THE ROLE OF PRODUCT ATTRIBUTES

Mohammad Rahman
Jungpil Hahn
Krannert School of Management
Purdue Univeristy

{mrahman, jphahn}@purdue.edu

Abstract

There is evidence of both increasing equality as well as inequality in the sales distribution on the Internet. This leads to the question of whether or not the features (tools) that are available on the Internet to lower search costs, e.g., search tools, detailed product information page, affect disparate product types differently. Consequently, the current study focuses on the following questions surrounding the changes in sales distribution that take on heightened importance for retailers and consumers: Do current system features homogenously affect the sales distribution of all product types? Can we generate insights regarding the impact of available tools that lowers search costs on the sales of different product types? Note that a product can be of two primary types: search goods and experience goods. Search goods refer to products that can be evaluated with direct product related information before purchase, e.g., a screwdriver. On the contrary, consumers can only correctly evaluate the fit of an experience good with their tastes after actually experiencing the product, such as by buying and subsequently eating gourmet cheese. Our results demonstrate that the sales distribution for search goods is significantly flatter than that of experience goods, a finding that underscores the importance of product attributes in determining the sales distribution on the Internet. We also show that the sales of search goods is insignificantly influenced by product rating, which signifies the decrease in a consumer's reliance on experience sharing mechanisms as she begins to gather more information that help her to asses the fit of the product with her tastes.

Keywords: Sales distribution, product attribute, Gini coefficient, Lorenz curve, product rating

1. Introduction

> "Overture and Google's success came from an understanding of what Chris Anderson refers to as "the long tail," the collective power of the small sites that make up the bulk of the web's content." (O'Reilly 2005)

The emerging argument regarding the sales of different products in the Internet economy, expressed in the preceding quote, is that the low search cost on the Internet will help niche products to account for relatively higher sales compared to what niche products can generate in traditional channels (Brynjolfsson et al. 2006). In the year 2000, back-listed books, in other words niche books which are not even stocked by typical conventional bookstores because of their insignificant sales, have accounted for about 40% of Amazon.com's book sales revenue (Brynjolfsson et al. 2003). This phenomenon calls for a change from the traditional business model, where a few popular products account for most sales, to a new model where niche items are pushed more to generate sales (Anderson 2004; Anderson 2006). However, it is not well-understood whether the Internet is uniformly shifting the balance; i.e., where popular products no longer account for most sales, for all product types.

Interestingly, recent research studies have found differing evidence regarding the shift in the sales distribution for different types of products. Brynjolfsson et al. (2006) has considered women's clothing and found a flatter sales distribution online. More specifically, when the products were ranked based on high sales to low sales and a curve was fitted to study the sales distribution on the Internet as well as on the Catalog channel, the slope was smaller for the Internet channel. This suggests that the tail of the sales distribution is longer on the Internet channel,--a phenomenon which is described as "the Long Tail" by Chris Anderson (Anderson 2004). In contrast, in a large experiment involving an artificial market, Salganik et al (2006) finds that hit songs are many times more successful than the rest, and the inequality in download distribution intensifies as consumers are allowed see the songs sorted in descending order based on the number of downloads (indicating the current popularity). In another study, Elberse and Oberholzer-Gee (2006) find that the number of non-selling video titles have increased in the time period 2002 to 2005. In addition, they find that among the best-performing titles, an ever-smaller number of video titles accounts for most sales, an indication of concentrated (steep) sales distribution. Thus, there is evidence of both increasing equality as well as inequality in the sales distribution on the Internet. This leads to the question of whether or not the features (tools) that are available on the Internet to lower search costs, e.g., search tools, detailed product information page, affect disparate product types differently.

Consequently, the current study focuses on the following questions surrounding the changes in sales distribution that take on heightened importance for retailers and consumers: Do current system features, such as search tools, detailed information, and product ratings, homogenously affect the sales distribution of all product types? Can we generate insights regarding the impact of available tools that lowers search costs on the sales of different product types? Subsequently, the objective of this research is to investigate, in the presence of same set of system features, whether or not the product attributes influence the sales distribution that is observed on the Internet. In particular, we shed light on the importance of the degree to which the current system features support the product attributes, which in turn, plays a significant role in shifting the balance of the sales distribution on the Internet.

To understand the shifts in sales distributions for different products, it is necessary to understand the product types and how the Internet affects them. A product can be of two primary types – search goods and experience goods – depending on whether or not the quality of the product may be assessed prior to consumption and use (Darby and Karni 1973; Nelson 1970; Nelson 1974). Search goods refer to products that can be evaluated with direct product related information before purchase, e.g., a screwdriver. Information cues allow a customer to correctly evaluate the search attributes of a product, thereby allowing the customer to find the fit cost ex-ante for search goods (Darby and Karni 1973; Nelson 1970; Nelson 1974).[1] In this regard, online markets are more likely to lower the uncertainty associated with evaluating search goods (Ford et al. 1990). On the contrary, consumers can only correctly evaluate the fit of an experience good with their tastes after actually experiencing the product, such as by buying and subsequently eating gourmet cheese (Darby and Karni 1973; Nelson 1970; Nelson 1974).

[1] Fit cost refers to the cost incurred by a customer if the product does not match her taste.

Information cues, such as detailed product information, seller's reputation, or product rating, may help a consumer to anchor the initial evaluation; notwithstanding, a consumer can only find the final fit cost for experience goods after experiencing the product. Since consumers face relatively greater degree of uncertainty in purchasing experience goods, popular products may be more favored in the presence of imperfect signals, such as those displayed by the herd behavior (Banerjee 1992; Bikhchandani et al. 1992; Shiller 1995).

Clearly, current system features, such as detailed product information, product rating, and search tools, theoretically make it easier for consumers to evaluate the search goods and address heterogeneity in taste. On the other hand, it is relatively more difficult to evaluate experience goods online.[2] Nonetheless, this has not yet been empirically validated. In this work, we empirically examine whether the current dominant designs of online markets, which theoretically favor a specific product type -- search goods -- are equally impacting the sales distribution of both product types.

Understanding how current system features have affected the sales distribution of different products has important managerial implications for multi-product firms. A retailer needs to understand the impact on the sales distribution of different product types in order to offer the right product variety as well as manage inventory. In addition, insights regarding the effects of current system designs enable a retailer to streamline the system features with product attributes. These, in turn, will enable retailers to exploit the changes in sales distribution to generate more profit. Subsequently, the answers to these questions have important implications as the Internet commerce now accounts for a large portion of the overall economy. To address these questions, we employ econometric methods to analyze the demands of a leading Internet book retailer.

Our results demonstrate that the sales distribution for search goods is significantly flatter than that of experience goods, a finding that underscores the importance of product attributes in determining the

[2] Nevertheless, it is likely that search tools, product ratings, and other system features make it easier to accommodate consumer heterogeneity for experience goods compared to the traditional channel. In this paper, we do not focus on comparing the sales distributions for each product type on the Internet channel and the traditional channel.

sales distribution on the Internet. We also show that the sales of search goods is insignificantly influenced by product rating, which signifies the decrease in a consumer's reliance on experience sharing mechanisms as she begin to gather more information that help her to asses the fit of the product with her tastes. We also focus on the impact of a rich search tool—Search Inside, which allows a consumer to search for a word within a book and/or read selected pages of a book—on the sales distribution for both product types.[3] We find strong evidence that the impact of such a tool is more pronounced on the search goods. Since a tool like Search Inside is consistent with the characteristics of search goods, it enables customers to relatively better asses the fit with their tastes.

The paper is organized as follows. The next section presents the theoretical background and develops the hypotheses for the study. Section 3 describes the methodology used in the paper. We present our empirical analyses in section 4. We discuss the theoretical and practical implications as well as limitations of the study in section 5.

2. Theory and Hypotheses

Bakos (1997; 1998) points out the efficiency of the electronic market in performing basic market functions, i.e. matching buyers and sellers, the facilitation of transactions, and institutional infrastructure. He highlights that because of relatively low search costs, consumers inquire about all available products and purchase the one that best fits their needs, leading to a socially optimal allocation. In this respect, online retailers or markets offer a wide variety of products that consumers probably wouldn't have heard of otherwise (Brynjolfsson et al. 2003). For instance, Amazon.com stocks many more books compared to a large conventional bookstore. Correspondingly, the welfare benefit from consumer's increased ability to search for and find a wide variety of books online is 5-7 times more important than the lower prices available on the Internet (Brynjolfsson et al. 2003).

---

[3] On the Amazon.com website, a subset of the books have Search/Look inside available. Search Inside allows a consumer to search for keywords and read a few pages relevant to the search. Similarly, Look Inside allows a consumer to read a selected portion of the book. For simplicity, we just use the word Search Inside if either one of these two tools are available for a book.

The impact of low search costs on price and price dispersion has been the subject of significant academic research. There is significant evidence that prices online are lower than their conventional counterparts (Brynjolfsson and Smith 2000; Goolsbee 2000), and there exists substantial price dispersion online (Brown and Goolsbee 2002; Brynjolfsson and Smith 2000; Clay et al. 2002; Clemons et al. 2002; Smith and Brynjolfsson 2001). However, the low search cost on the Internet may not only affect the price competition, but also influence what products generate sales on the Internet. On the Internet, if a consumer finds and purchases a product that fits her tastes which is otherwise hard to find in traditional stores, then the Internet is likely to flatten the sales distribution on the Internet. This is because rather than only a few SKUs accounting for all sales (like the traditional channel), relatively more SKUs will be credited for sales on the Internet. Not surprisingly, the impact of low search costs on the concentration of product sales online has been gaining more attentions in recent research (e.g., Brynjolfsson et al. 2006; Elberse and Oberholzer-Gee 2006).

Brynjolfsson et al. (2006) demonstrates that the sales distribution online is significantly flatter than the sales distribution on the catalog channel. In other words, they find strong evidence of the "Long Tail" effect online. However, they do not focus on investigating whether the impact of low search costs on sales distribution is different for different types of products. As mentioned earlier, search goods can benefit substantially from information cues, whereas the uncertainty associated with experience goods are not eliminated to the same extant with information cues.

The theory of economics of information posits predicts that a consumer's uncertainty about search goods decreases with greater amount of verifiable information cues. On the other hand, a consumer's uncertainty about experience goods decreases after sampling or experiencing the products (Ford et al. 1990). Therefore, the information contents available online reduces fit costs for search goods more than it does for experience goods (Klein 1998). In other words, the low search costs on the Internet allow consumers to relatively efficiently search for, find, and evaluate products with search attributes that fit their tastes. Hence, because of the fit between search cost lowering features (e.g., search tools, information page) and product characteristics the Internet is likely to radically impact the sales distribution for search goods.

On the other hand, the current characteristics of the Internet are not designed to eliminate the uncertainty associated with experience goods (Ford et al. 1988; Jain et al. 1995; Jin et al. 2005).[4] Although a consumer can benefit from product reviews that are available online to screen products (Chevalier and Mayzlin 2006), she cannot asses the fit of an experience good with her taste until experiencing it. Thus, the degree to which the current features that are available online can support the attributes of experience goods in order to eliminate uncertainty is very limited. Hence, the impact of the Internet, in terms of flattening the sales distribution, on the sales distribution of experience goods is likely to be less pronounced than that on the sales distribution of search goods. In other words, we expect to see a flatter sales distribution for search goods compared to the sales distribution for experience goods.

H1: *The sales distribution of search goods will be flatter than the sales distribution of experience goods on the Internet.*

To understand the sales of individual products, it is important to consider the search behavior of a consumer. Consumers undertake more searches for search goods as opposed to experience goods. Consequently, in the presence of low search costs, consumers are likely to gather more information cues about search goods rather than relying on the recommendation of others. In contrast, in the case of experience goods, consumers demonstrate greater reliance on recommendation by others, e.g., product ratings. As such, consumers rely heavily on their own experience or others' experiences for experience goods (Klein 1998). As a result, product rating is likely to have relatively less influence on the sales of search goods as compared to the influence on the sales of experience goods.

H2: *The influence of product rating will be higher on the sales of experience goods compared to the sales of search goods.*

---

[4] The uncertainties associate with experience goods can be reduced by the means of product demonstration or sample distribution (Jain et al. 1995; Klein 1998). Typically, a consumer can asses the fit of such a product with her taste after experiencing it.

The widely prevalent search tools and product information pages enables a consumer to find a product and gather some information cues about the product. In this vein, a rich search tool that allows gathering content related information cues can help a consumer to further reduce uncertainties regarding the fit with the consumer's taste. Therefore, the sales distribution is likely to be relatively flatter for both the search goods and the experience goods in the presence of a rich search tool. Once again, since the fit with the experience goods can only be evaluated post consumption, the impact of such a rich tool on the sales distribution of search goods is likely to be more pronounced than that on the sales distribution of experience goods

H3: *A rich search tool will flatten the sales distribution for both product types.*

As mentioned earlier, if consumers are able to gather information cues that help them asses the fit of a product at low cost, then they tend to rely less on experiences shared by others about the product. Note that traditionally on the Internet, it has been difficult for consumers to have direct access to (part of) the product prior to purchase. Consequently, as argued earlier, traditional features have been mostly helping consumers to relatively better asses the fit with search goods, which are likely to reduce a consumer's reliance on the experiences shared by others. However, with the introduction of rich search tools that provides content related information, consumers are positioned to further reduce the uncertainty associated with the fit of an experience good as well as a search good. As a result, it is likely that consumers will rely less on others' recommendations in the presence of a rich search tool compared to the case where the rich search tool does not exist. Consequently, the relative effect of product ratings on sales for both product types is likely to decrease in the presence of a rich search tool.

H4: *In case of both product types, the effect of product rating on sales will be lower in the presence of a rich search tool.*

Next, we subject the theoretical prediction that search goods are more likely to have flatter sales distribution than experience goods on the Internet to empirical investigation, and show the importance of product attributes in analyzing impact of the Internet on the product sales distribution. We also focus on testing the rest of the hypotheses and provide insights in the coming sections.

3. Methodology

In this study, it is critical to capture the reaction of the consumers to the current electronic market design in natural settings. Accordingly, we conduct a field study in the book industry to test the predictions. The field study gives us the opportunity to capture the real consumer purchase behavior in the book industry. We picked the book industry since this is one of the industries where both product types -- search and experience goods – are present. Furthermore, 40% of Amazon's book sales revenue in 2000 has been attributed to obscure books, which are not even offered for sale by conventional retailers (Brynjolfsson et al. 2003).[5]

For this study, we consider reference books as search goods since, given enough description, most buyers can assess the fit of the book with their tastes. On the other hand, we consider fictions as experience goods since, in most cases, only after reading fiction books, are readers able to evaluate whether they liked it or not.

3.1 Data Collection

We collected our demand data from Amazon.com, a major online book retailer, which adequately represents overall online demand for books. Moreover, Amazon is a pioneer in developing emerging features, such as Search Inside. It is important to note that Amazon.com provides similar search tools, product information, rating mechanism, etc. for both the reference books and fictions. Thus, the tools that are available to lower search costs are identical for both product categories.

Since there are a large number of fiction titles available on Amazon.com, we have randomly selected 100 page title listings, which generated 5314 fictions. Similarly, we picked 2450 reference books from Amazon.com. We wrote web crawlers (programs) to collect the data from Amazon.com website. For every single book we recorded its price, sales rank, rating, number of raters, date of publication, and

[5] Thus, we have picked an industry where the "Long Tail" effect has been widely documented at least for some categories of books, which, in turn, is biased against our test of disparate impact on the sales distribution of the experience goods compare to that of the search goods. As a result, if we find support for differential effect on the sales distributions of search goods and experience goods, it will be a conservative estimate. In other words, such a design strengthens the findings of our study.

whether search/look-inside is available between March 6 and April 6, 2006. Our web crawler has collected the relevant information for each book at the same time during the day during the data collection period.

In our sample, there are books that were unavailable for purchase from Amazon, although other information was available. We have excluded such books from our analysis as there cannot be any sales for them at Amazon.com, although other retailers may sale them at Amazon Marketplace. Note that Amazon sales rank does not consider any sales from Amazon Marketplace. In addition, the Search Inside status for 11 reference books has changed during our study period. Similarly, the Search Inside status has changed for 72 fictions between March 6, 2006 and April 6, 2006. In order to eliminate the effect of such change on the sales, we included only the books for which the Search Inside status remained constant during study period.

Our final sample contains 3798 fictions and 1333 reference books. Table 1 presents the descriptive statistics for both fictions and reference books. As we can see from the table, there were 2352 fictions with Search Inside feature. Correspondingly, there were 537 reference books with Search Inside feature.

Since Amazon.com does not report actual demand data, we use Pareto relationship to infer the actual weekly demands from the reported ordinal sales rank. This Pareto relationship: $Quantity = \delta.Rank^{\beta}$ has been widely used in literature to infer weekly demands (e.g., Brynjolfsson et al. 2003; Chevalier and Goolsbee 2003; Ghose et al. 2006).

In order to mitigate the effect of any sudden shift in Amazon sales rank, we have aggregated the sales rank for each book over one month and calculated the average sales rank for each book. Following past studies (i.e. Brynjolfsson et al. 2003; Chevalier and Goolsbee 2003; Ghose et al. 2006), we have estimated the demand from the following relationship with the coefficient from Brynjolfsson et al. 2003:

$$\log(Quantity) = \beta_1 + \beta_2 . \log(Rank), \text{ where } \beta_1 = 10.526 \text{ and } \beta_2 = -0.871.$$ [6]

4. Results

4.1. Descriptive Results

Table 1 shows that on an average 0.45 reference book had been sold weekly, while fictions' weekly sales had been on an average 1.54. Although a few fictions generated a large amount of sales (maximum is 244.05), the range of sales for both book categories were very similar (only 12 fictions had sales more than maximum sales of reference books (40.84)). Within each book category, the range of sales for books with Search Inside and without Search Inside was also very similar. A Fiction with Search Inside had an average sale of 1.94 while the average sale for a fiction without Search Inside was 0.89. Correspondingly, the average sale for a reference book with Search Inside was 0.72 while the average sale for a reference book without Search Inside was 0.26. The average Amazon price of a fiction was $20.69 while the average Amazon price of a reference book was $16.10. On an average 58.67 raters rated fictions, and the average rating of a fiction was 4.33. On the other hand, on an average 3.63 people rated reference books, and the average rating for a reference book was 4.45. A fiction received about 14% discount while a reference received about 10% discount from list price.

Table 1: Descriptive Statistics

	Reference Books	Fictions
Avg(Sales)	0.45	1.54
Max(Sales)	40.84	244.05
Min(Sales)	0.062	0.062
Avg(Amazon price)	$16.10	$20.69
Avg(Discount)	9.65%	13.86%
Avg(Rating)	4.45	4.33
Avg(Number of Raters)	58.67	3.63
w/Search Inside	537	2352
Total	1333	3798

[6] We have also used the estimate reported by Chevalier and Goolsbee (2003) in order to check the robustness of our results. Their estimated $\beta_2 = -0.855$. Reassuringly, the qualitative nature of the results remains the same. In accordance with most extant studies, we have reported the results using the estimates from Brynjolfsson et al. 2003.

4.2. Analyses

The Lorenz curves (Lorenz 1905) and the Gini coefficients (Gini 1921) have been widely used in the literature to characterize the inequality in wealth and income distribution (e.g., Petersen 1979). Brynjolfsson et al. (2006) have first used Lorenz curves and Gini coefficients to study the sales distribution online. The Gini coefficient is a measure of distributional inequality, a number between 0 and 1, where 0 corresponds with perfect equality. In particular, 0 Gini coefficient embodies that all the products in a category have identical demand, and 1 corresponds with perfect inequality, or one product has all the demand, and all other books have zero demand.

The Gini coefficient is based on the Lorenz curve, a commonly used representation of distributional equality, most commonly used to compare income distributions across regions and time. The Lorenz curve for a product category can be derived by ranking all the products in increasing order of sales, and then by plotting the cumulative share $F(\theta)$ of sales associated with each ascending rank percentile θ, where $0 < \theta \leq 1$. Complete equality, where sales are shared equally among all products within a category, is represented by a 45-degree straight line (see Figure 1). Therefore, the closer the Lorenz curve is to the straight line, the more equal is the sales distribution within a product category. The Gini coefficient is computed as twice the area between the Lorenz curve and the 45-degree line between the origin and (100%, 100%).

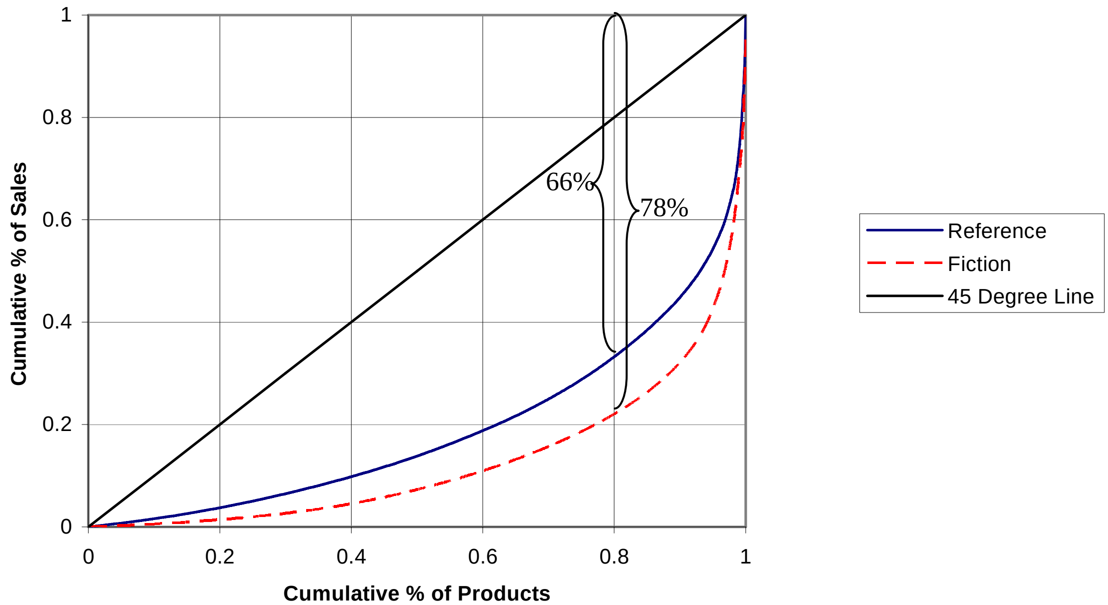

Figure 1: Lorenz Curves (Gini coefficients: Refrence books 0.62; Fictions 0.75)

Notably, the Gini coefficient measures inequality in the demand distribution, regardless of the category's average demand (or popularity), which facilitates comparing fictions and reference books despite their intrinsic differences, even if exists any (The World Bank 2007).

Accordingly, we have constructed the Lorenz curves and computed Gini coefficients for both fictions and reference books using all the books in respective categories. The Lorenz curves are presented in Figure 1. The solid curve represents the sales distribution for reference books and the dashed curve represents the sales distribution for fictions. The Lorenz curve for reference books lies above the fiction books' Lorenz curve and closer to the 45 degree line, indicating that the sales distribution is more equal for the reference books than the fiction books. Correspondingly, the Gini coefficient (0.62) for reference books is lower than that (0.75) for fictions.

It is evident from the Lorenz curves that the top 20% of the reference books account for about 66% of sales, whereas the top 20 % of fictions account for about 78% of overall sales. The dominance of top 20% books in fiction category remarkably resembles the Pareto rule (i.e., 80/20 rule). In contrast, the sales distribution for reference books is more equal, as expected. Accordingly, we find evidence from the Gini coefficients that the sales distributions of fictions and reference books on the Internet are different. Although the Gini coefficients are widely used as a measure of inequality of wealth distribution, due to the lack of standard error estimate, we cannot measure whether the difference between the fictions and the reference books are statistically significant. In such situations, a widely used method is to use bootstrapping to empirically construct distributions that can be statistically compared (Efron and Gong 1983). In addition, bootstrapping has been found to be an effective method to estimate statistical significance of Gini coefficients (Modarres and Gastwirth 2006). Accordingly, we have constructed a bootstrap sample of 500 for each category and calculated the Gini coefficients 10,000 times for each category -- fictions and reference books.

By the law of large numbers, we can assume normality for both samples and conduct a two-sample t-test to see whether the mean Gini coefficients are different. We find that the mean Gini coefficient for reference books is 0.613 and the mean Gini coefficient for fictions is 0.737. The t-test rejects the null hypothesis that the means are equal for both book categories (t-statistic = 214.95, $p < 0.001$). Consequently, the results suggest that the Gini coefficient for fictions is significantly different and smaller than reference books' Gini coefficient.[7] Thus, we find support for the first hypothesis that the sales distribution for search goods is flatter than the sales distribution for experience goods.

---

[7] A Jarque-Bera normality test (Jarque and Bera 1980; Jarque and Bera 1987) rejects the null hypothesis that the Gini coefficients are from a normal distribution in both bootstrapped distributions. We find that the median Gini coefficient for reference books is 0.614 and the median Gini coefficient for fictions is 0.736. The non-parametric Mann-Whitney rank test strongly rejects the null hypothesis that the median coefficients for both categories are the same (z-value = 119.78, $p < 0.001$). Reassuringly, we find strong evidence that the Gini coefficient for reference books is different than the Gini coefficient for fictions.

Table 2: The Effect of Product Rating on Sales

	Sales
AmazonPrice	-0.021**
	(0.004)
ProductRating	-0.095
	(0.065)
NumberOfRaters	0.011**
	(0.003)
Discount	1.556**
	(0.544)
FictionDummy	-0.997
	(0.556)
ProductRating×FictionDummy	0.321**
	(0.120)
Intercept	1.223**
	(0.333)
Observations	4056
F-statistic	13.26
R^2	0.059

Robust Standard errors are listed in parentheses.
Significance: $^{*}\, p < 0.05$, $^{**}\, p < 0.01$

Table 3: Gini Coefficients – Grouped by with or without Search Inside

	With Search Inside	Without Search Inside
Reference Books	0.65	0.48
Fictions	0.74	0.72

Next, we focus on testing the second hypothesis, which relates to the influence of product rating on sales. This can be tested by estimating the effect of product rating on the sales of both fictions and reference books. Note that the price of a product has significant influence on the sales of a product. Similarly, the discount associated with a product influences the demand of the product. Finally, the total number of raters can have influence on the demand as well as influence the overall rating. Therefore, we estimate the effect of product rating on the sales of both fictions and reference books, after controlling for the price, discount, number of raters, and the inherent differences between the two book categories. Accordingly, we estimate the following regression model:

$$\begin{aligned} Sales = \beta_0 + \beta_1 AmazonPrice + \beta_2 ProductRating + \beta_3 NumberofRaters + \beta_4 Discount \\ + \beta_5 FictionDummy + \beta_6 ProductRating * FictionDummy + \varepsilon \end{aligned}$$

Here, *Discount* refers to the price discount a book received from the list price. The *FictionDummy* takes the value 1 for all fictions and zero otherwise. Table 2 presents the estimates of the model. We can see from the estimates that as the price increases, the sales of a book decrease (*AmazonPrice;* β_1 = -0.021, $p < 0.01$). As the discount offered for a book price increases, the sales increase (*Discount;* β_4 = 1.556, $p < 0.01$). The coefficient associated with the *FictionDummy* is not significantly different from zero (β_5 = -0.997, n.s.), which suggests that there is perhaps no intrinsic difference in sales for these two categories. For testing H2, we are interested in the estimates of β_2 and β_6. Note that the coefficient estimate for *ProductRating* is not significantly different from zero (β_2 = -0.095, n.s.), which suggests that the sales of reference books are not significantly influenced by the product rating. In contrast, the coefficient estimate for the interaction between product rating and fiction dummy is positive and significantly different than zero (β_6 = 0.321, $p < 0.01$). This suggests that as the product rating increases, sales of fictions also increase. Thus, we find support that product rating has significant positive influence on the sales of fictions, whereas it has insignificant influence on the sales of reference books. Thus, H2 is supported.

To test the third hypothesis, which relates to affect of a rich search tool on the sales distribution, we analyze the impact of a rich search tool -- Search Inside -- on the sales distribution of both fictions and reference books. As described earlier, Search Inside is available for a subset of books in both categories. Thus, a comparison of the impact of Search Inside on the sales distribution within the category positions us to test H3. Table 3 presents the Gini coefficients for both categories, each separated into two groups – with Search Inside and without Search Inside.

We find that the Gini coefficient for reference books with Search Inside is 0.65, whereas the Gini coefficient for reference books without Search Inside is 0.48. In contrast, the Gini coefficient for fictions with Search Inside is 0.74, whereas the Gini coefficient for fictions without Search Inside is 0.72.

Once again, because of the lack of a standard error estimate, we cannot measure whether the difference between different groups are statistically significant. Subsequently, we have constructed empirical distribution for all four groups using bootstrapping to study whether or not the difference in Gini coefficients is statistically significant. We find that the mean (0.73) Gini coefficient for fictions with Search Inside is significantly different from the mean (0.71) Gini coefficient for fictions without Search Inside (t-statistic = -37.11, $p < 0.01$). Correspondingly, the mean (0.64) Gini coefficient for reference books with Search Inside is significantly different from the mean (0.47) Gini coefficient for reference books without Search Inside (t-statistic = -206.73, $p < 0.01$).[8] Since the Gini coefficients for books with Search Inside are higher for both categories, H3 is not supported.

Finally, to test the fourth hypothesis, we need to estimate the effect of product rating within each category by differentiating the books that have Search Inside from those without Search Inside. We estimate the effect of product rating on the sales within each category, after controlling for the price, discount, number of raters, and the presence/absence of Search Inside. Accordingly, we estimate following regression model for each book category:

$$\begin{aligned} Sales = \beta_0 + \beta_1 AmazonPrice + \beta_2 ProductRating + \beta_3 NumberofRaters + \beta_4 Discount \\ + \beta_5 SearchInsideDummy + \beta_6 ProductRating * SearchInsideDummy + \varepsilon \end{aligned}$$

Here, the SearchInsideDummy takes the value 1 if the feature is available for a book, zero otherwise. Table 4 presents the estimates for both fictions and reference books. For testing H4, we are interested in the estimates of β_2 and β_6. We find that the coefficient associated with product rating is significantly different than zero for fictions (*ProductRating;* $\beta_2 = 0.550$, $p < 0.01$), but not reference books (*ProductRating;* $\beta_2 = 0.068$, n.s.). This is consistent with the earlier finding in that product rating, in general, has insignificant impact on the sales of reference books but has significant positive influence on the sales of fictions. However, the coefficient for the interaction effect is significantly different from zero and negative for reference books ($\beta_6 = -0.195$, $p < 0.05$). Hence, we find support that, in the presence of

[8] The Jarque-Bera normality test (Jarque and Bera 1980; Jarque and Bera 1987) rejects the null hypothesis that the Gini coefficients are from a normal distribution in all four bootstrapped distributions. Nonetheless, the Mann-Whitney rank test reaffirms the results that are reported in the text. The test statistic for reference books is: z-value = -117.18, $p < 0.001$. Correspondingly, the test statistic for fictions is: z-value = 31.69, $p < 0.001$.

Search Inside, product rating has relatively less influence on the sales of reference books. Similarly, the coefficient for the interaction effect is significantly different from zero and negative for fictions (β_6 = -0.467, $p < 0.05$). This suggests that the impact of product rating is lower on the sales of fictions with Search Inside. Thus, H4 is supported.

Table 4: The Effect of Product Rating on Sales in the Presence of Search Inside

	Reference Books	Fictions
AmazonPrice	-0.002	-0.020**
	(0.005)	(0.003)
ProductRating	0.068	0.550**
	(0.044)	(0.157)
NumberOfRaters	0.148*	0.011**
	(0.062)	(0.003)
Discount	1.530**	1.143
	(0.552)	(0.641)
SearchInsideDummy	1.170*	3.148**
	(0.454)	(0.900)
ProductRating SearchInsideDummy*	-0.195*	-0.467*
	(0.094)	(0.202)
Intercept	-0.434	-1.869**
	(0.340)	(0.720)
Observations	614	3442
F-statistic	3.17	13.59
R^2	0.155	0.063

Robust Standard errors are listed in parentheses.
Significance: * $p < 0.05$, ** $p < 0.01$

4.3. Robustness Check

As mentioned earlier, the Gini coefficient considers any inherent differences that may exist between product categories. Nevertheless, the robustness of the finding that the impact of low search cost on the Internet on the sales distribution is contingent on the product attributes can be checked by further analyzing the impact of Search Inside on the sales distribution of different book categories. Note that if the current features available on the Internet, that lowers the search cost, affects all product types similarly, then the impact of Search Inside should not be radical on the sales distribution of reference books compare to the impact of Search Inside on the sales distribution of fictions.

In order to compare the effect of Search Inside on the sales distribution of fictions and reference books, we conduct a two-way ANOVA of the Gini coefficients generated from the bootstrapping (recall that empirical distributions of Gini coefficients were generated for all four groups in order to test H2). The ANOVA reinforces the result that the sales distributions for reference books are flatter than the sales distributions for fictions in both cases – with Search Inside and without Search Inside (F-statistic = 105,433.55, $p < 0.01$).

More importantly, we are interested in whether or not the difference between the mean Gini coefficient (0.64) for reference books with Search Inside and mean Gini coefficient (0.47) for reference books without Search Inside is larger than the difference between the mean Gini coefficient (0.73) for fictions with Search Inside and mean Gini coefficient (0.71) for fictions without Search Inside. Figure 2 presents line graph of the interaction effect. Correspondingly, the change in the Gini coefficients for reference books is significantly different than the change in the Gini coefficients for fictions (F-statistic = 18,703.56, $p < 0.01$). This suggests that the impact of Search Inside is more pronounced on the sales distribution of reference books compare to the impact of Search Inside on the sales distribution of fictions.

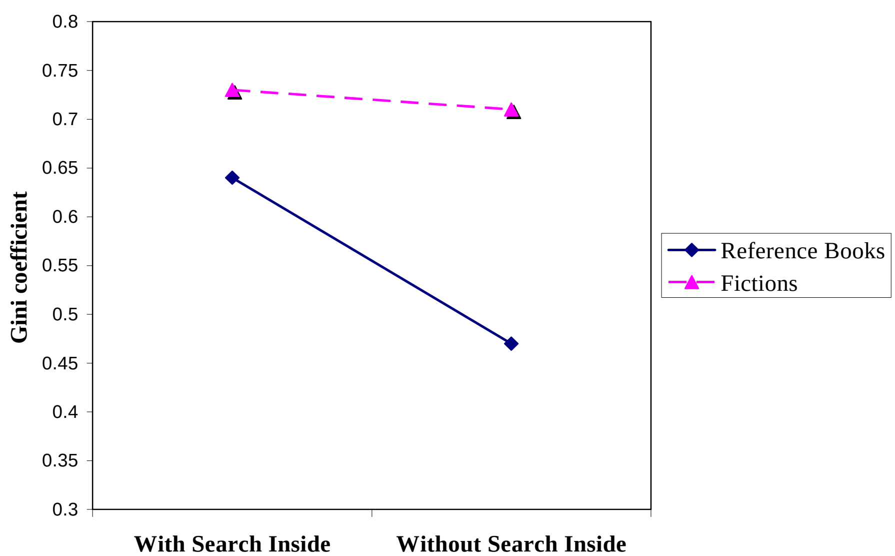

Figure 2: Interaction Graph

5. Discussion and Conclusion

The results of this study have significant theoretical and practical implications. On the theoretical front, our results suggest that it is important to consider the product attributes in analyzing the impact of low search cost on the Internet on the sales distribution of products. In particular, we find that the sales distribution of search goods is more influenced by the features available, e.g., search tools, on the Internet. The sales distribution of search goods is flatter than the sales distribution of experience goods on the Internet. This demonstrates that technology enabled information cues help consumers to relatively better asses the fit of search goods with their tastes compare to assessing the fit of experience goods with their tastes. This is because a consumer cannot evaluate the fit of an experience good with her taste until experiencing the product.

In addition, we find evidence that as product rating increases the sales for experience goods increase, whereas product rating has smaller and insignificant correlation with the sales of search goods. Not surprisingly, consumers rely heavily on the experience of others in evaluating experience goods. On the other hand, since the Internet allows a consumer to gather information cues at a low cost and enables her to remove relatively more uncertainty associated with search goods, the impact of product rating is negligible on the sales of search goods. Also, we show that with the introduction of richer search tools that can further eliminate the uncertainty associated with product evaluation, the impact of product rating on the sales of experience goods decreases.

These findings add to the emerging literature on the impact of the Internet on the sales distribution by underscoring the significance of the degree to which the tools available on the Internet fit various product attributes. Our study also has important implications for the nascent literature on the impact of word-of-mouth, product rating, etc. on the sales of different products. We demonstrate that experience sharing mechanisms have differential effect on the sales for different product types. Therefore, researchers need to be careful about the product attributes and how various system features complement the product attributes in studying the impact of such experience sharing mechanisms.

With the increasing importance of the Internet in the overall economy, our study has important implications for multi-product retailers. A retailer needs to understand the impact of the Internet on the sales distribution of different product types in order to offer the right product variety as well as manage inventory. Our results show that a retailer may expect to sell more niche products in case of search goods; however, the sales distribution for experience goods is likely to be relatively dominated by popular products in the current setup. Also, due to the relatively high degree of uncertainty that is associated with experience goods, consumer may prefer to purchase this type of products from physical store where they are relatively better positioned to examine such a product. Forman et al. (2006) have found that consumers prefer to buy fictions from local stores, whereas they tend to buy non-fictions from the Internet. Therefore, managers need to be careful in choosing the product mix they offer on the Internet. In choosing the right product mix, practitioners need to consider to what degree the tools (features) available on the Internet support different product attributes.

As mentioned earlier, our results demonstrate that in the presence of system features that help a consumer to eliminate uncertainty associated with the fit of the product with her taste, the importance of experience sharing, e.g., product rating, in increasing sales decreases. This implies that practitioners need to be strategic about using experience sharing mechanisms for different products. Uniformly utilizing experience sharing mechanisms for all products, regardless of the product attributes, may not allow a retailer to reap the full benefit of such mechanisms.

Interestingly, we find that the sales distribution for products with the Search Inside is significantly steeper than that without the Search Inside. This raises interesting new possibilities. One potential explanation could be that such a tool allows consumer to gather information about a product not only along the horizontal dimension but also along the vertical (quality) dimension. The extant studies heretofore have considered the impact of the Internet on the sales distribution by assuming that products are horizontally differentiated. In reality, the products sold on the Internet are likely to be horizontally as well as vertically differentiated. Correspondingly, a rich search tool that allows consumer to determine the quality of the product prior to purchase should have an impact on sales distribution. Once again, since such a tool reduces search cost significantly but does not allow a consumer to experience the product, this tool is likely to affect search goods more significantly.

If all the search goods are of the same quality or consumers' quality preferences are horizontally differentiated, then the sales distribution should be even flatter in the presence of rich search tools compare to the one in the presence of traditional search tools. On the contrary, if products are vertically differentiated and consumers' quality preferences are homogeneous, then high quality products will account for all the sales, ceteris paribus. Although rigorously investigating the possible impact of gaining information about the quality of a product with Search Inside is beyond the scope of this paper, we do see that the average product rating is lower for books with Search Inside—4.39 for reference books and 4.32 for fictions, whereas the average product rating for books without Search Inside is 4.52 for reference books and 4.34 for fictions. The lower average product rating for books with Search Inside strengthens the conjecture that high quality books account for a significant portion of overall sales, thereby concentrates the sales distribution.

Future research should investigate whether system features, e.g., search tools, product information page, enables a consumer to evaluate the product along the horizontal dimension and/or along the vertical dimension, and what is the subsequent impact on the sales distribution of such features. Note that understanding how each feature individually affects the sales distribution of different product types can be valuable in making decision regarding whether or not make a feature available for a particular product type. It would be interesting to study the incremental effect of introducing various features on the sales distribution of different types of products. It would also be beneficial to understand how different system features change the sales distribution for different types of products compare the traditional channel, where such features are absent.

One of the limitations of our data is that we infer actual demand from sales rank. Also, future studies need to focus on other industries in order to generalize the results of the current study. Due to the lack of data we cannot identify how does different system features impact the sales distribution of different types of products.

6. References

Anderson, C. "The Long Tail " in: *Wired Magazine*, 2004.

Anderson, C. *The Long Tail: Why the Future of Business Is Selling Less of More*, (1st ed.) Hyperion, New York, 2006, p. 256.

Bakos, J.Y. "Reducing buyer search costs: Implications for electronic marketplaces," *Management Science* (43:12) 1997, p 1676.

Bakos, Y. "The Emerging Role of Electronic Marketplaces on the Internet," *Communications of the ACM* (41:8) 1998, pp 35-42.

Banerjee, A. "A simple model of herd behavior," *Quarterly Journal of Economics* (107) 1992, pp 797-818.

Bikhchandani, S., Hirshleifer, D., and Welch, I. "A Theory of Fads, Fashion, Custom, and Cultural-Change as Informational Cascades," *Journal of Political Economy* (100:5), Oct 1992, pp 992-1026.

Brown, J.R., and Goolsbee, A. "Does the Internet make markets more competitive? Evidence from the life insurance industry," *Journal of Political Economy* (110:3), Jun 2002, pp 481-507.

Brynjolfsson, E., Hu, Y.J., and Simester, D. (2006). "Goodbye Pareto Principle, Hello Long Tail: The Effect of Search Costs on the Concentration of Product Sales." SSRN, from http://ssrn.com/paper=953587

Brynjolfsson, E., and Smith, M.D. "Frictionless Commerce? A Comparison of Internet and Conventional Retailers," *Management Science* (46:4) 2000, p 563.

Brynjolfsson, E., Yu, H., and Smith, M.D. "Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers," *Management Science* (49:11) 2003, pp 1580-1596.

Chevalier, J., and Goolsbee, A. "Measuring Prices and Price Competition Online: Amazon.com and BarnesandNoble.com," *Quantitative Marketing and Economics* (V1:2) 2003, pp 203-222.

Chevalier, J.A., and Mayzlin, D. "The effect of word of mouth on sales: Online book reviews," *Journal of Marketing Research* (43:3), Aug 2006, pp 345-354.

Clay, K., Krishnan, R., Wolff, E., and Fernandes, D. "Retail strategies on the Web: Price and non-price competition in the online book industry," *Journal of Industrial Economics* (50:3), Sep 2002, pp 351-367.

Clemons, E.K., Hann, I.H., and Hitt, L.M. "Price dispersion and differentiation in online travel: An empirical investigation," *Management Science* (48:4), Apr 2002, pp 534-549.

Darby, M.R., and Karni, E. "Free Competition and the Optimal Amount of Fraud," *Journal of Law & Economics* (16:1) 1973, pp 67-88.

Efron, B., and Gong, G. "A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation," *American Statistician* (37:1) 1983, pp 36-48.

Elberse, A., and Oberholzer-Gee, F. "Superstars and Underdogs: An Examination of the Long Tail Phenomenon in Video Sales," Harvard Business School Working Paper Series, 2006, pp. 1-50.

Ford, G.T., Smith, D.B., and Swasy, J.L. "An Empirical Test of the Search, Experience and Credence Attributes Framework," *Advances in Consumer Research* (15:1) 1988, pp 239-243.

Ford, G.T., Smith, D.B., and Swasy, J.L. "Consumer Skepticism of Advertising Claims: Testing Hypotheses from Economics of Information," *Journal of Consumer Research* (16:4) 1990, p 433.

Forman, C., Ghose, A., and Goldfarb, A. (2006). "Geography and Electronic Commerce: Measuring Convenience, Selection, and Price." from http://ssrn.com/paper=941175

Ghose, A., Smith, M., and Telang, R. "Internet Exchanges for Used Books: An Empirical Analysis of Product Cannibalization and Welfare Impact," *Information Systems Research* (17:1) 2006, pp 3-19.

Gini, C. "Measurement of Inequality of Incomes," *The Economic Journal* (31:121) 1921, pp 124-126.

Goolsbee, A. "In A World without Borders: The Impact of Taxes on Internet Commerce," *Quarterly Journal of Economics* (115:2) 2000, pp 561-576.

Jain, D., Mahajan, V., and Muller, E. "An Approach for Determining Optimal Product Sampling for the Diffusion of a New Product," *Journal of Product Innovation Management* (12:2) 1995, pp 124-135.

Jarque, C.M., and Bera, A.K. "Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals," *Economics Letters* (6:3) 1980, pp 255-259.

Jarque, C.M., and Bera, A.K. "A Test for Normality of Observations and Regression Residuals," *International Statistical Review* (55:2), Aug 1987, pp 163-172.

Jin, Y., Zilberman, D., and Heiman, A. "Fit Risk: Secondhand Market versus Money-back Guarantee," University of California at Berkeley, 2005.

Klein, L.R. "Evaluating the Potential of Interactive Media through a New Lens: Search versus Experience Goods," *Journal of Business Research* (41:3) 1998, pp 195-203.

Lorenz, M.O. "Methods of measuring the concentration of wealth," *Publications of the American Statistical Association* (9) 1905, pp 209-219.

Modarres, R., and Gastwirth, J.L. "A cautionary note on estimating the standard error of the gini index of inequality," *Oxford Bulletin of Economics and Statistics* (68:3), Jun 2006, pp 385-390.

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

Nelson, P. "Advertising as Information," *Journal of Political Economy* (82:4) 1974, p 729.

O'Reilly, T. "What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software," O'Reilly, http://www.oreilly.com, 2005.

Petersen, H.G. "Effects of Growing Incomes on Classified Income Distributions, the Derived Lorentz Curves, and Gini Indexes," *Econometrica* (47:1) 1979, pp 183-198.

Salganik, M.J., Dodds, P.S., and Watts, D.J. "Experimental study of inequality and unpredictability in an artificial cultural market," *Science* (311:5762), Feb 2006, pp 854-856.

Shiller, R.J. "Conversation, information, and herd behavior," *American Economic Review* (85:2) 1995, p 181.

Smith, M.D., and Brynjolfsson, E. "Consumer Decision-making at an Internet Shopbot: Brand Still Matters," *Journal of Industrial Economics* (49:4) 2001, p 541.

The World Bank "Measuring Inequality," http://go.worldbank.org/3SLYUTVY00, 2007.