

Goodbye Pareto Principle, Hello Long Tail: The Effect of Search Costs on the Concentration of Product Sales

This version: November 2007

Erik Brynjolfsson*, Yu (Jeffrey) Hu **, Duncan Simester***

ABSTRACT

Many markets have historically been dominated by a small number of best-selling products. The Pareto Principle, also known as the 80/20 rule, describes this common pattern of sales concentration. However, by greatly lowering search costs, information technology in general and Internet markets in particular have the potential to substantially increase the collective share of niche products, thereby creating a longer tail in the distribution of sales.

This paper investigates how demand-side factors contribute to the Internet's "Long Tail" phenomenon. It first models how a reduction in search costs will affect the concentration in product sales. Then, by analyzing data collected from a multi-channel retailing company, it provides empirical evidence that the Internet channel exhibits a significantly less concentrated sales distribution, when compared with traditional channels. The difference in the sales distribution is highly significant, even after controlling for consumer differences. Furthermore, the effect is particularly strong for individuals with more prior experience using the Internet channel. We find evidence that Internet purchases made by consumers with prior Internet experience are more skewed toward obscure products, compared with consumers who have no such experience. We observe the opposite outcome when comparing purchases by the same consumers through the catalog channel. If the relationships we uncover persist, the underlying trends in technology and search costs portend an ongoing shift in the distribution of product sales.

We thank Chris Anderson, Jerry Hausman, Lorin Hitt, Hank Lucas, Jiwoong Shin, Hal Varian, Pai-Ling Yin, seminar participants at Carnegie Mellon University, New York University, University of Minnesota, University of Pennsylvania, Yale University, the Workshop on Information Systems and Economics (WISE), and the 2007 American Economic Association Annual Meeting (AEA) for valuable comments on this paper. Generous funding was provided by the MIT Center for Digital Business and NSF Grant IIS-0085725.

* MIT Sloan School of Management; email: erikb@mit.edu; web: <http://digital.mit.edu/erik>
** Purdue University, Krannert School of Management; email: yuhu@purdue.edu
*** MIT Sloan School of Management; email: simester@mit.edu

1. Introduction

Many markets have traditionally been dominated by a few best-selling products. Book sales are concentrated on a few books by several established best-selling authors (Greco 1997); Billboard “top 40” hits account for the lion’s share of radio playlists and music sales; and movie rentals are dominated by a few “new releases”. Economists and business managers often use the Pareto Principle to describe this phenomenon of sales concentration. The Pareto Principle, which is sometimes called the 80/20 rule, states that a small proportion (e.g., 20 percent) of products in a market often generate a large proportion (e.g., 80 percent) of sales.¹ However, the Internet has the potential to shift this balance. Anderson (2004) has coined a term—“The Long Tail”—to describe the phenomenon that niche products make up a large share of total sales. On the Internet, the Pareto Principle may be giving way to the “Long Tail”.

Anecdotal evidence suggests that Internet markets has helped shift the balance of power from a few best-selling products to niche products that are previously obscure. For example, Frank Urbanowski, Director of MIT Press, observes that the increased accessibility to backlist titles through the Internet has resulted in a 12% increase in sales of these titles (Professional Publishing Report 1999). This increase has happened despite flat growth in overall book sales. Similar observations have been made in electronic markets for music, DVDs, and electronics. Rhapsody, an online music provider, streams more songs each month beyond its top 10,000 than it does its top 10,000. While “new release” movies account for a dominant share of revenue in a video rental shop, DVDStation, a company that allows consumers to search and reserve movies online and pick them up in a DVD kiosk, finds that more than 50% of their rental revenue come from titles that are not new releases (DVDStation 2005).

Previous research has focused on *supply-side* explanations for the Internet’s Long Tail phenomenon. For example, Brynjolfsson, Hu and Smith (2003) document how centralized inventory and drop-shipping agreements allow online book retailers to offer over 2 million book titles (and millions more used and out-of-print titles). In contrast, physical space restrictions, logistics, and holding costs limit the product

¹ The 80/20 rule was first suggested by Vilfredo Pareto in his study of wealth distribution (Pareto 1896), and has since been applied to the analysis of city population, product sales, and sales force management.

selection in a typical brick-and-mortar book store to between 40,000 and 100,000 titles. In this paper we reverse the lens and investigate how *demand-side* factors contribute to the Internet's Long Tail. More specifically, Internet markets provide consumers with search tools, browsing tools, and recommendation systems. If these Internet tools facilitate consumer search, then lower consumer search costs may have an impact on the concentration of product sales, leading to the Long Tail phenomenon on the Internet.

Typically both supply-side and demand-side explanations exist concurrently. For instance, a firm can respond to lower consumer search costs (a change in the demand side) by increasing its product selection (a change in the supply side). Therefore, it can be difficult to fully disentangle these explanations and attribute the change in the concentration of product sales to only supply-side factors or to only demand-side factors. However, a unique feature of our dataset is the ability to distinguish demand-side and supply-sided explanations. The data was provided by a retailer that sells through both Internet and catalog channels. The two channels use the same order fulfillment methods and facilities. This controls for differences in sales tax policies, shipping costs, product selection, stockouts, and other supply-side factors. Most importantly, because of this common fulfillment structure, customers can order an identical selection of products and at an identical set of prices through both channels. However, these two channels have important differences. The Internet channel provides online search and recommendation tools that help customers locate niche products, while the catalog channel does not have these tools. Thus, supply-side factors are held constant, while demand-side factors differ across the channels.

This paper studies the effect of search costs on the concentration of product sales. We provide empirical evidence that the Internet channel, when compared with the catalog channel, exhibits a significantly less concentrated sales distribution. To make comparisons, we introduce the use of the Gini Coefficient and Lorenz Curve, concepts that are traditionally used to study the inequality in income and wealth distribution (Lorenz 1905, Gini 1912), to measure the concentration of product sales. We find that the difference in the sales distribution across these two channels is highly significant, even after controlling for consumer differences. Furthermore, the effect is particularly strong for individual consumers with

more prior experience using the Internet channel. We find evidence that Internet purchases made by consumers with more prior Internet experience are more skewed toward niche products, compared with consumers who have little or no such experience. Because more-experienced users presumably are more adept at searching, we interpret this result as further evidence of the effect of search costs on the concentration of product sales.

The primary focus of the paper is empirical, but the paper also provides a simple theoretical model of consumer search behavior. In this model a retailer can reduce search costs on a sample of “key” products, while consumers have to incur search costs to find the “niche” products. The purpose of this theoretical model is to illustrate that a reduction in the search cost for niche products can lead to a decrease in the concentration of product sales. This model helps motivate the empirical work that follows.

Previous Literature

Understanding how lower search costs on the Internet can affect the concentration of product sales has important managerial implications for a firm that sells multiple products to consumers. By lowering consumer search costs, Internet markets may expand the set of products consumers consider when making their purchases. Consumers are no longer limited to “key” products that are heavily promoted, advertised and, as a result, highly visible. The increased sales of niche products can lead to changes in the firm’s profit function and changes in the firm’s strategies in producing and marketing its products.

Remarkably, despite the abundance of anecdotal evidence and the important managerial implications, theoretical and empirical research on how search costs affect the concentration of product sales is virtually non-existent. Theoretical research has provided predictions on how search costs can affect price, price dispersion, entry, and product variety (e.g., Diamond 1971, Wolinsky 1986, Anderson and Renault 1999, Bakos 1999, Cachon, Terwiesch, and Xu 2006). But these models assume a homogeneous search cost for all products. As a result, a reduction in search costs affects all products symmetrically, and do not change the concentration of product sales. Our model allows for heterogeneity in search costs across products. As a result, a reduction in search costs affects the products asymmetrically, leading to a change in the concentration of product sales.

There is also a growing body of empirical research studying how a reduction in buyer search costs on the Internet can impact prices and price dispersion (e.g., Brynjolfsson and Smith 2000, Morton, Zettelmeyer, and Silva-Risso 2001, Brown and Goolsbee 2002, Hann, Clemons and Hitt 2003, Clay, Krishnan, Wolff, and Fernandes 2003). However, these studies do not consider how a variation in search costs affects the concentration of product sales.

Recently, researchers have begun to study the Long Tail phenomenon on the Internet. Ghose and Gu (2006) use aggregate-level demand data to estimate how search costs can affect market competition. Our paper differs from, and therefore complements, their paper by using individual-level transaction data and by focusing on the effect of search costs on the concentration of product sales, instead of the effect on market competition. Elberse and Oberholzer-Gee (2007) study the distribution of aggregate-level video sales. They find that sales become concentrated on a smaller number of titles among best-performing titles and that the number of less-popular titles that sell only a few copies a week has increased from 2000 to 2005. Their paper does not distinguishing demand-side and supply-side explanations of the Long Tail phenomenon, while our paper focuses on demand-side explanations and studies this problem using individual-level demand data.

Structure of the Paper

In Section 2 we begin with a model illustrating how search costs affect the concentration of product sales. Section 3 describes the design of our empirical analyses. In Section 4, we present the empirical findings and the paper concludes in Section 5 with a review of the findings and broader implications.

2. Model

In this section, we provide a theoretical model of consumer search behavior. The results derived in this model are used to motivate our empirical work that is the main focus of this paper.

Products

We consider a multi-product monopolist selling n horizontally differentiated products to consumers. The firm's n products include $k \in (1, n)$ “key” products and $n-k$ “niche” products. The key products are

prominently displayed so that all customers can easily find them. Examples of key products include products that are displayed at the entrance, by the cash register, at the end of the aisle, and products that appear on the most visible catalog pages or Web pages. Because of their high visibility, consumer search costs are low for key products. In contrast, consumers have to exert effort and incur search costs in order to find niche products. Specifically, we assume key products have zero search cost, and niche products have a positive search cost s .

We focus on the study of demand-side explanations of the Long Tail phenomenon on the Internet and we hold the supply-side factors constant. Thus, the number of available products (n) and the number of key products (k) are held constant. The variable cost of production (c) is assumed to be the same across all n products.

Consumers

There is a population of L consumers with heterogeneous tastes. Each consumer $i=1, 2, \dots, L$ has a utility function of the form:

$$u_{ij} = v_{ij} - p_j, \quad j=1, 2, \dots, n, \quad (1)$$

where $(v_{i1}, v_{i2}, \dots, v_{in})$ are the values consumer i attaches to the n products, and $(p_1, p_2, \dots, p_n)$ are the prices for the n products. We assume that v_{ij} is a random variable that is independently and identically distributed. Thus, the heterogeneity in consumer tastes is captured by the difference in realized values of $(v_{i1}, v_{i2}, \dots, v_{in})$ for different consumers.²

A consumer can learn about the values she places on the n available products only by searching. A consumer's search is without replacement and with recall. That is, each time a consumer searches, she

² This form of product differentiation with random consumer utilities has now been widely used in the economic literature. It is first introduced by Perloff and Salop (1985) and later adopted by Wolinsky (1986), Anderson and Renault (1999), and others. It represents an improvement from earlier literature that uses a straight line or a unit circle to model product differentiation.

learns about a different product and the value she places on the product. She can proceed to purchase any one of the products she has already searched without incurring any additional cost.³

Each time a consumer searches, she learns about a different product. With probability $\alpha \in (0,1)$, this product does not fit her taste; with probability $1 - \alpha$, this product fits her taste. The value consumer i places on product j is zero if product j does not fit her taste ($v_{ij} = 0$), and is equal to t_i if product j fits her taste does ($v_{ij} = t_i$).⁴ We assume t_i is heterogeneous across consumers and is distributed according to a Uniform distribution with support on $[0,1]$. Each consumer has unit demand, and so after purchasing one unit of any of the n products, she exits the market.

Notice that the firm will prefer to charge the same price for all key products, and we use p_K to denote this price. Similarly, it will charge the same price for all niche products, which we denote by p_N . We assume that these prices are observed by consumers before search. Each consumer also knows the number of available products (n), the number of key products (k), the search cost for niche products (s), the probability that a product she finds after each search does not fit her taste (α), and the value she places on a product that fits her taste (t_i). However, before searching a consumer does not know the values she places on each product (v_{ij}); the value consumer i places on product j is revealed only after consumer i has searched and evaluated product j .

Consumers' Searching and Purchasing Strategies

Because key products have zero search costs, a consumer will always search all of the k key products and find the product that gives her the highest value. When there are multiple products that give consumer i the same highest value, she will choose one product randomly from these products to represent the key product with the highest value. Let w_i be the value that consumer i places on this product. Because w_i is

³ This specification of search without replacement and with recall is consistent with Wolinsky (1986) and Anderson and Renault (1999). Alternatively, an assumption of search with replacement will lead to an optimum search policy and expected outcome that is arbitrarily close to the optimum search policy and expected outcome under this specification, as long as the horizon of search is long enough.

⁴ This specification of a product either fits a consumer's taste or does not fit a consumer's taste simplifies the model and allows us to obtain a close-form solution. Similar assumptions have been used in Schulz and Stahl (1996), which studies how consumer search affects the equilibrium price in a market with differentiated products.

the maximum of k random variables that are independently and identically distributed according to a Bernoulli distribution with failure probability α , we know that $w_i = 0$ with probability α^k , and $w_i = t_i$ with probability $1 - \alpha^k$. Having observed the value of w_i , a consumer has three options: exiting the market without purchasing; exiting the market after purchasing the key product that gives her the highest value amongst all key products; or continuing to search in the hope of finding a niche product that gives her a higher utility (a lower price, a higher value or both) than the key product that gives her the highest net value. Because the benefit from searching niche products is a function of w_i , α , t_i , p_K , and p_N , a consumer's decision to search depends upon these variables. The following claim describes the searching and purchasing strategy that maximizes a consumer's expected utility.

Claim 1

1. When none of the key products fit consumer i 's taste ($w_i = 0$)

If $s \geq (1 - \alpha)(1 - p_N)$ consumer i will not search niche products and instead will exit the market

without purchasing. If $0 \leq s < (1 - \alpha)(1 - p_N)$ and $t_i > p_N + \frac{s}{1 - \alpha}$, consumer i searches and

purchases the first niche product that fits her taste. If no such product is found after searching through all the niche products, she will exit the market without purchasing. If $0 \leq s < (1 - \alpha)(1 - p_N)$ and

$t_i \leq p_N + \frac{s}{1 - \alpha}$, consumer i will not search and instead will exit the market without purchasing.

2. When one of the key products fits consumer i 's taste ($w_i = t_i$) and $p_N \geq p_K$

Consumer i will not search niche products. If $t_i > p_K$ consumer i will purchase the key product that gives her the highest value; otherwise she will exit the market without purchasing.

3. When one of the key products fits consumer i 's taste ($w_i = t_i$) and $p_N < p_K$

If $s \geq (1-\alpha)(p_K - p_N)$, consumer i will not search niche products. As long as $t_i > p_K$ she will purchase the key product that gives her the highest value; otherwise she will exit the market without purchasing. If $0 \leq s < (1-\alpha)(p_K - p_N)$ and $t_i > p_K$ then consumer i searches and purchases the first niche product that fits her taste. If no such niche product is found, she will purchase the key product that offers the highest value. If $0 \leq s < (1-\alpha)(p_K - p_N)$ and $p_N + \frac{s}{1-\alpha} < t_i \leq p_K$, consumer i also searches and purchases the first niche product that fits her taste. If none of the niche products fit her taste she exits the market without purchasing. If $0 \leq s < (1-\alpha)(p_K - p_N)$ and $t_i \leq p_N + \frac{s}{1-\alpha}$ consumer i will not search and instead will exit the market without purchasing.

A complete proof of all the Claims and Propositions can be found in the Appendix. Notice that if one additional search yields a negative net benefit, then the net benefit from subsequent search opportunities is also negative. This result, first proved by Kohn and Shavell (1974), explains why consumers only focus on one additional search and do not consider the value of subsequent search opportunities. In equilibrium the firm chooses prices to maximize its profits, while consumers optimize their search strategies. Claim 2 characterizes the equilibrium demand and profits.

Claim 2

The equilibrium profit and demands are given by the following expressions:

	High Search Costs $s > (1 - \alpha)(1 - c)$	Low Search Costs $s \leq (1 - \alpha)(1 - c)$
Key Product Demand $D_F^*(s, n, k)$	$\frac{L}{2k}(1 - c)(1 - \alpha^k)$	$\frac{L}{2k}(1 - c)(1 - \alpha^k)$
Niche Product Demand $D_N^*(s, n, k)$	0	$\frac{L}{2(n - k)} \left(1 - \frac{s}{1 - \alpha} - c \right) (\alpha^k - \alpha^n)$
Equilibrium Profit $\pi^*(s, n, k)$	$\frac{L}{4}(1 - c)^2(1 - \alpha^k)$	$\frac{L}{4}(1 - c)^2(1 - \alpha^k) + \frac{L}{4} \left(1 - \frac{s}{1 - \alpha} - c \right)^2 (\alpha^k - \alpha^n)$

In equilibrium the firm sets the prices so that customers who find a suitable key product do not search niche products. Therefore, expected demand for the key products is unaffected by the cost of searching niche products. However, expected demand for the niche products increases when search costs are lower, and so expected profits are also higher (notice that $\alpha^k > \alpha^n, \forall k < n$). We restate this result as a formal proposition.

Proposition 1. Equilibrium profits decrease when consumer search costs for niche products are higher.

That is, $\frac{\partial \pi^*(s, n, k)}{\partial s} \leq 0$ for any $s \geq 0$.

Next we consider the effect of search costs on the concentration of product sales. Economists have long used the Lorenz Curve and Gini Coefficient to describe the inequality in income (wealth) distribution (Lorenz 1905, Gini 1912), although we are not aware of any previous research that has applied these two concepts to the study of sales distributions. The Lorenz Curve is drawn inside a square box with the x-axis measuring the cumulative percentage of individuals and the y-axis measuring the cumulative percentage of income (wealth) held by the individuals. The Gini Coefficient is the ratio of the area between a Lorenz Curve and a 45 degree line to the total area under a 45 degree line. Formally, it is defined as:

$$G = 1 - \frac{2 \sum_{j=1}^n [(n+1-j)d_j]}{(n+1) \sum_{j=1}^n d_j}, \quad (2)$$

where the demand for n products are denoted by $(d_1, d_2, \dots, d_n)$ and $d_1 \leq d_2 \leq \dots \leq d_n$.

When wealth is perfectly evenly distributed across individuals, the Lorenz Curve coincides with a 45 degree line and the Gini Coefficient equals zero. As the distribution becomes more concentrated, the Lorenz Curve curves away from a 45 degree line and the Gini Coefficient increases. In this paper, we will apply the Gini Coefficient to measure sales concentration. We derive the following proposition characterizing the effect of search costs on the Gini Coefficient.

Proposition 2. Product sales become more concentrated when consumer search costs for niche products are higher. That is, $\frac{\partial G(s, n, k)}{\partial s} \geq 0$ for any $s \geq 0$.

In equilibrium, a consumer who does not find any suitable key products search the niche products if and only if the value she places on a product that fits her taste (t_i) exceeds a threshold. As search costs decrease, this threshold declines and more consumers search niche products. As a result, niche products generate a larger share of total sales and we observe a less concentrated sales distribution.

Extension: Consumers with Different Search Costs

So far we have assumed that all consumers in the model have the same search costs. This is consistent with prior theoretical research that has studied how search costs can affect price, price dispersion, entry, and product variety (e.g., Diamond 1971, Wolinsky 1986, Bakos 1999, Anderson and Renault 1999, Cachon, Terwiesch, and Xu 2006). However, it is possible that past experience or expertise may result in some consumers having lower search costs than others. Next, we will consider the variation in search costs across consumers and study how a consumer's search costs can affect her tendency to purchase niche products.

We assume the firm still sells the same set of k key products and $n-k$ niche products to consumers. The firm now sells these products at the same set of prices to two samples of consumers with different search costs for niche products. There are L consumers in Sample 1 whose search costs for niche products are s_1 , while there are L consumers in Sample 2 whose search costs for niche products are s_2 . We assume $s_1 > s_2$. As long as Sample 1 and Sample 2 consumers receive the same set of prices, consumers with higher search costs are more likely to restrict their purchases to key products. At the same time, the threshold for searching is lower for Sample 2 consumers who have lower search costs, and so they are more likely to purchase niche products. As a result, niche products generate a larger share of total sales for Sample 2 consumers and we observe a less concentrated sales distribution for Sample 2 consumers. We formally prove this result as a proposition.

Proposition 3. Sample 2 consumers, whose search costs are lower, have a higher tendency to purchase niche products; product sales are less concentrated for Sample 2 consumers than for Sample 1 consumers.

That is, $D_N(p_K, p_n, s_2, n, k) \geq D_N(p_K, p_n, s_1, n, k)$, $G(p_K, p_n, s_2, n, k) \leq G(p_K, p_n, s_1, n, k)$ for any $s_1 > s_2$ and for any p_K and p_n .

Summary

In equilibrium, a *decrease* in search costs for niche items unambiguously leads to a *decrease* in the concentration of product sales. Variations in search costs may occur across retail channels or across consumers. In our empirical analysis we will test Proposition 2 and Proposition 3 by exploiting both sources of variation. In the next section we describe the context for our empirical analyses, and we then present findings in Section 4.

3. Design of Empirical Analyses

The company we study is a medium-sized retailer selling mainly women's clothing at moderate price levels.⁵ All of the products carry the company's private label brand and are sold exclusively through the company's catalog channels (mail and telephone), Internet website and a small number of physical retail

⁵ The company asked to remain anonymous.

stores. We have data describing purchases by consumers from 1998, when the company first started selling over the Internet, through 2002. Because the company was unable to adequately identify the relatively small number of consumers purchasing in its brick-and-mortar stores, we do not have a record of purchases made by consumers in this channel. This limits our study to the Internet and catalog channels.

A key feature of the company is that it offers the same product selection and prices through its Internet and catalog channels. This policy simplifies the company's logistics and ordering processes. In addition, it avoids potential consumer dissatisfaction if consumers observe that they have paid higher prices for an item than other consumers (see for example Anderson and Simester 2007). There are differences in search costs across the two channels, due in part to specific technologies like recommendation engines and search tools available only via the Internet. However, the company does not respond by varying the number of available products across the channels.⁶

In addition to offering the same products through both channels, the company also uses the same order fulfillment methods and facilities for the two channels. This controls for differences in sales tax policies, shipping costs, and the possibility of stockouts, eliminating several alternative explanations for potential differences in the sales distribution across the two channels.

Recall our distinction between "supply-side" and "demand-side" drivers of the Long Tail. Supply-side explanations refer to shelf-space and other logistical restrictions, such as the cost of holding inventory, and/or the cost of shipping and coordinating the distribution of products to multiple stores. Demand-side explanations include features that relate to the cost of customer search, even when the features are *supplied* by the retailer. These include differences in the availability of customer search engines, and limitations on the number of products featured on web pages or in retail catalogs. Accordingly, the data we study vary in demand-side drivers, but not supply-side drivers.

⁶ In 2002, 9,110 unique products registered positive sales through either the catalog channel or the Internet channel. Among them, 9,074 products (99.6%) registered positive sales through the catalog channel and 6,247 products (68.6%) registered positive sales through the Internet channel. The percentage of products that registered positive sales is smaller for the Internet channel, partly because the Internet channel accounted for a smaller percentage of total sales, compared with the catalog channel, in 2002.

We distinguish between the catalog and Internet channels by the manner in which customers place their orders. Customers ordering through the catalog channel place their orders either by calling the company's 1-800 number and speaking to a telephone service representative, or by completing the physical order-form bound into the middle of a catalog and mailing this to the company. Approximately 90% of orders through the catalog channel are made by telephone. Orders through the Internet channel are made by customers visiting the company's website and using its online ordering system. The percentage of orders placed through this channel was initially small, but grew gradually. By 2002 Internet sales accounted for approximately 10% of orders (see later discussion for additional details).

There are differences across the channels in the process used to feature key products. In the catalog channel key products are featured by including them in the catalog (not all of the product range appears in each catalog), together with the page location and size of the presentation. Products that appear on the inside or outside covers tend to receive the most attention, as do products within the catalog located nearer the front page. Featuring key products on the Internet channel also involves product location. The company decides which products appear on the "Home" page and the first page of each successive category.

Variations in Search Costs across the Channels

Search costs vary across the two channels. Products that are featured in a catalog can be located by consumers with relative ease, but consumers have to make significant efforts to locate a non-featured product, by going through all the pages in a current catalog. With a little more effort, consumers can dig up catalogs sent out earlier and purchase products in those catalogs. Some consumers even engage in long conversations with sales representatives over the phone to identify products and make purchases. However, the Internet channel greatly facilitates consumer searches by allowing consumers to search by product keyword and item number. It also provides an online recommender system, which affects the ease with which products can be located. These active and passive tools dramatically reduce consumer search costs (Hoque and Lohse 1999, Brynjolfsson, Hu and Smith 2006).

We will exploit the variation in search costs across the two channels and compare the concentration of product sales across the Internet and catalog channels. Our research approach is consistent with previous research that has compared Internet and offline channels (e.g., Brynjolfsson and Smith 2000, Morton, Zettelmeyer, and Silva-Risso 2001, Brown and Goolsbee 2002, Hann, Clemons and Hitt 2003, Clay, Krishnan, Wolff, and Fernandes 2003).

We recognize that search costs and the numbers of key products may co-vary across the channels. Thus, when we compare the concentration of product sales across the Internet and catalog channels, we do not seek to distinguish whether differences in the sales distribution are due to differences in the number of key products in each channel or differences in search costs. Indeed, given the nature of search activities on the Internet it is not clear that it is meaningful to think of these two constructs as operating independently. If the firm invests in an improved recommendation tool, should this be interpreted as lowering search costs, or increasing the number of featured products? A similar question arises if the firm designs a search tool so that it highlights a larger number of products when customers search on a keyword.

Variations in Search Costs across Consumers

Search costs also vary across consumers, providing a second way to assess our model. Hann and Terwiesch (2003) find that prior Internet experience can reduce consumers' frictional costs in online transactions, one type of search costs. As consumers gain more experience using online search tools and recommender systems at the company's site, they become more familiar with these systems and can use them more effectively. Thus their search costs become lower. A consumer who has extensive prior experience purchasing through the company's Internet channel is more likely to be familiar with the site's searching, browsing, and recommendation features and have a lower search cost than a consumer with little no prior experience purchasing from the Internet channel. We will use an individual-level analysis to study whether consumers' prior Internet experience is correlated with their tendency to purchase niche

products. We note that such an analysis focuses on a single channel—the Internet, and as a result, is not confounded by differences in the process used to feature key products across the two channels.

We begin by presenting results comparing the concentration of sales across the two channels. We will then focus on a single channel and compare the concentration of sales across consumers within that channel.

4. Empirical Results

Our empirical analysis focuses on three sets of questions:

1. Does the Internet channel exhibit a less concentrated sales distribution? (Proposition 2)
2. Does the difference in the concentration of product sales across the two channels survive if we control for consumer differences by focusing on a common sample of consumers?
3. How does a consumer's experience with the Internet channel affect her tendency to purchase niche products? (Proposition 3)

We begin by constructing Lorenz Curves and Gini Coefficients for the two channels. We then control for consumer differences by restricting our attention to consumers who have purchased through both channels. Finally, we measure the extent to which prior Internet experience moderates these outcomes.

Initial Results

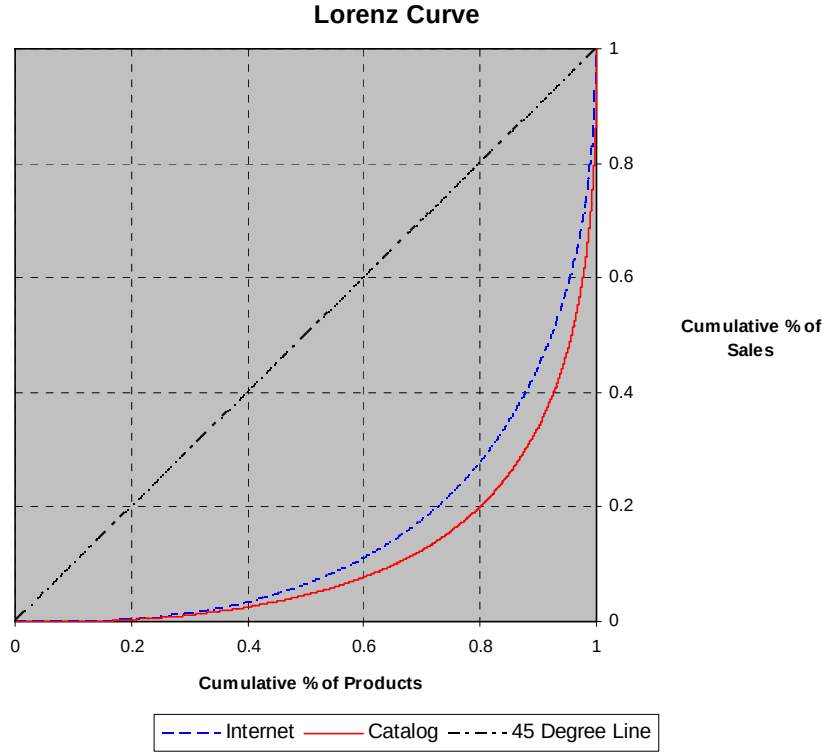
Prior to 2002 there were relatively few Internet transactions and so most of our analyses focus on transactions that occurred in 2002. We will later use the earlier data to control for customer differences. During 2002 the company sold 7,725,574 items through the catalog channel, and 702,659 items through the Internet channel. In comparing the concentration of sales we want to ensure that the findings are not affected by the differences in aggregate demand. Therefore, we randomly select 100,000 items purchased from each channel to represent demand in that channel. Overall, these 100,000 items include 5,568 unique products in the Internet channel, and 5,595 products in the Catalog channel. The union of these two samples yields 6,550 unique products purchased through both channels, with an intersection of 4,613

products. We construct Lorenz Curves and calculated Gini Coefficients for both channels using all 6,550 products. Reassuringly, our results are robust to either using all of the transactions from each channel, or randomly selecting the same number of products from each channel.

The Lorenz Curves are presented in Figure 1. The blue (dashed) curve represents the Internet channel and the red (solid) curve represents the catalog channel. The Lorenz Curve for the Internet channel lies above the catalog channel's Lorenz Curve, indicating that the sales distribution is less concentrated for the Internet channel than for the catalog channel. Correspondingly, the Gini Coefficient for the Internet channel (0.70) is lower than that for the catalog channel (0.77). For the catalog channel, the bottom 80% of products generates 20.1% of sales. Interestingly, this corresponds with remarkable precision to the classic 80/20 Pareto Principle. In contrast, for the Internet channel, the bottom 80% of products generates 27.7% of sales.⁷

⁷ Economists have also used the ratio between the 95th percentile and the 50th percentile as a measurement for income inequality (e.g. Garicano and Hubbard 2006). The ratio between the 95th percentile and the 50th percentile is 13.75 for the catalog channel and 9 for the Internet channel, again indicating greater concentration in the catalog channel..

Figure 1: Lorenz Curve and Gini Coefficient for the Internet and Catalog Channels



Internet Gini Coefficient: 0.70; Catalog Gini Coefficient: 0.77

The Lorenz Curves and Gini Coefficients both suggest that there is a difference between the Internet and catalog sales distributions. However, these two tools do not allow us to conclude whether this difference is statistically significant. In order to do so, we fit the sales and sales rank data to a log-linear relationship and compare the coefficient on sales rank that is obtained when using data from the two channels. In particular we estimate the following log-linear relationship:

$$\ln(\text{Sales}_j) = \beta_0 + \beta_1 \ln(\text{Sales Rank}_j) + \varepsilon_j. \quad (3)$$

The log-linear curve described by Equation (3) is also known as Pareto Curve. Previous research has shown that it fits the relationship between product sales and sales rank very well across the full distribution of products (Brynjolfsson, Hu and Smith 2003, Chevalier and Goolsbee 2003). This curve has also been widely used by economists to successfully describe the distribution of income, wealth, and city size (Pareto 1896, Zipf 1949). Given this specification, β_0 can be interpreted as a measure of the

overall demand in the channel, while β_I measures how quickly the share of channel demand attributed to product j falls as the sales rank increases. Our prediction that lower search costs on the Internet are likely to result in a “long-tail” of sales distribution suggests that β_I will be less negative in the Internet channel than in the catalog channel. That is, in the Internet channel, products that have large sales rank numbers will retain a larger share of channel demand.

We estimate Equation (3) separately for the Internet and catalog demand data and report both sets of findings in Table 1. When estimating the model we again focus on a random sample of 100,000 items purchased from each channel. Because the dependent measure is only defined for products that have positive demand, we randomly selected 4,000 products that had positive demand in the Internet channel and 4,000 products that had positive demand in the catalog channel. The findings in Table 1 (and in the tables that follow) are again robust to alternative approaches, including: using all of the transactions; using all the products with positive demand in each channel; or using a fixed sample of the most popular products in each channel.

For the Internet channel, both coefficients are significantly different from zero, while the high R^2 value suggests that the log-linear relationship fits the data well. Our focus is on the comparison of the β_I coefficients between the two models. The difference in these coefficients is highly significant ($p < 0.01$),⁸ indicating that the distribution of product sales is significantly less concentrated in the Internet channel than in the catalog channel.

⁸ The t-statistic of this difference is:

$$\begin{aligned} \frac{(\beta_{1I} - \beta_{1C})}{\sqrt{\text{Var}(\beta_{1I} - \beta_{1C})}} &= \frac{(\beta_{1I} - \beta_{1C})}{\sqrt{\text{Var}(\beta_{1I}) + \text{Var}(\beta_{1C}) - 2\text{Cov}(\beta_{1I}, \beta_{1C})}} \\ &\geq \frac{(\beta_{1I} - \beta_{1C})}{\sqrt{\text{Var}(\beta_{1I}) + \text{Var}(\beta_{1C}) + 2\sqrt{\text{Var}(\beta_{1I})}\sqrt{\text{Var}(\beta_{1C})}}} \\ &= 7.8 \end{aligned}$$

Table 1: Regression of Sales onto Sales Rank

	Internet	Catalog
Constant	10.751** (0.058)	11.315** (0.048)
Log (Sales Rank)	-1.141** (0.008)	-1.250** (0.006)
R ²	0.852	0.909
Sample Size	4,000	4,000

Standard errors are in parentheses.

**Significantly different from zero, $p < 0.01$.

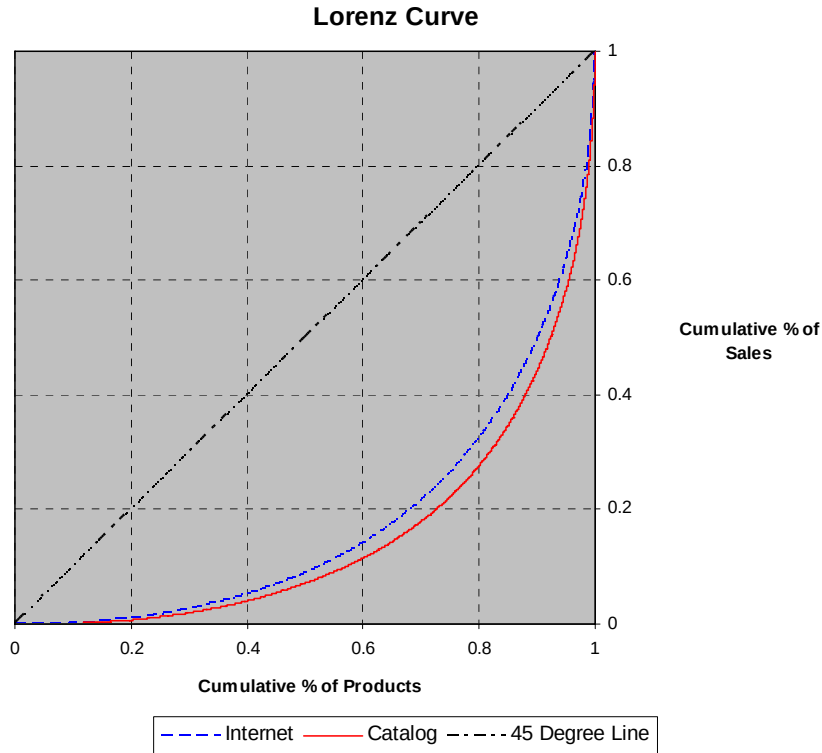
*Significantly different from zero, $p < 0.05$.

Controlling for Consumer Differences

There are factors other than lower search costs that can lead to a less concentrated sales distribution on the Internet. Our explanation is that search costs are lower on the Internet, while a possible alternative explanation is that consumers who shop through the Internet channel may be systematically different from consumers who shop through the catalog channel. Our data provides an opportunity to evaluate this alternative explanation. A total of 53,752 consumers purchased at least one item through each channel during 2002. By focusing on these consumers, we can compare the concentration of product sales across these two channels on a common sample of consumers.

Figure 2 presents two Lorenz Curves constructed using this common sample of consumers. To control for differences in aggregate demand, we again focus on a random sample of 100,000 items purchased from each channel. Overall, these 100,000 items included 5,626 unique products in the Internet channel, and 5,474 products in the Catalog channel. Their union yields 6,044 unique products purchased through both channels, with an intersection of 5,056 products. The Lorenz Curves and Gini coefficients are both calculated using the superset of 6,044 unique products represented in these 200,000 transactions. The Internet channel's Lorenz Curve lies above the catalog channel's Lorenz Curve, implying that the sales distribution is less concentrated for the Internet channel than for the catalog channel. Correspondingly, the Gini Coefficient for the Internet channel (0.65) is lower than that for the catalog channel (0.70).

**Figure 2: Lorenz Curve and Gini Coefficient for the Internet and Catalog Channels
Controlling for Consumer Selection**



Internet Gini Coefficient: 0.65; Catalog Gini Coefficient: 0.70

We also use the transaction data from this common sample of consumers to re-estimate Equation (3). We again randomly select 4,000 products that had positive demand in the Internet channel and 4,000 products that had positive demand in the catalog channel. The findings are reported in Table 2. They replicate the earlier results: the sales distribution has a longer tail in the Internet channel even when holding the sample of consumers fixed.⁹

⁹ The t-statistic of the difference in β_I equals 4.4:.

**Table 2: Regression of Sales onto Sales Rank
Controlling for Consumer Selection**

	Internet	Catalog
Constant	10.587** (0.059)	10.970** (0.057)
Log (Sales Rank)	-1.111** (0.008)	-1.177** (0.007)
R ²	0.839	0.865
Sample Size	4,000	4,000

Standard errors are in parentheses.

**Significantly different from zero, $p < 0.01$.

*Significantly different from zero, $p < 0.05$.

Notice that the difference in the Gini Coefficients (and the β_i coefficients) is smaller when we restrict our attention to a common sample of consumers. One interpretation for the attenuation in these differences is that there are differences in the consumers purchasing through the two channels, and this consumer difference contributes to the difference in the concentration of product sales across the channels. However, this is only a partial explanation for this result, as the difference in the concentration of product sales survives even when using a common sample of consumers.

Differences in Search Costs across Consumers

Next we will exploit the variation in search costs across consumers. Search costs on the Internet may depend upon a consumer's relative experience with the company's Internet channel (e.g., Hann and Terwiesch 2003). A consumer who has extensive prior experience is more likely to be familiar with the site's searching, browsing, and recommendation features. This familiarity may lower customer's effective search costs, allowing them to search more quickly through alternative product offers. According to Proposition 3, we would expect to find that the Internet purchases made by consumers with prior Internet experience are skewed towards niche products.

We develop two measures of a consumer's prior Internet experience. First, we define a dummy variable *Prior Internet Experience_i* as equal to one if consumer i had Internet experience prior to January 1, 2002; and zero otherwise. We also calculate the number of days from a consumer's first Internet purchase to

January 1, 2002, and use this as an alternative measure of a consumer's prior Internet experience (we label this measure *Days Since First Internet Purchase_i*).

For each consumer we identify all of the products purchased through the Internet channel in 2002 and calculate the average sales rank of these products (ranking the products according to their aggregate Internet demand). We use this *Average Sales Rank_i* variable as a measure of how consumer *i*'s Internet purchases are skewed towards niche products, with a higher *Average Sales Rank_i* indicating that consumer *i*'s purchases are more skewed towards niche products.

We then regress the natural log of *Average Sales Rank_i* onto a constant and the dummy variable *Prior Internet Experience_i*. Under this specification, the coefficient on the dummy variable *Prior Internet Experience_i* compares the percentage change in *Average Sales Rank_i* for consumers with and without prior Internet experience. We also include explicit controls for the heterogeneity in consumers' purchasing characteristics. In particular, we use the *Recency_i*, *Frequency_i*, and *Monetary Value_i* of consumers' historical transactions in the years prior to January 1, 2002.¹⁰ These so-called "RFM" measures, widely used in the catalog industry and in the marketing literature to segment consumers, provide natural candidates for control variables for consumer heterogeneity. This yields the following model specifications:

$$\ln(Average\ Sales\ Rank_i) = \alpha + \beta_1 Prior\ Internet\ Experience_i + \beta_2 \ln(Recency_i) + \beta_3 \ln(Frequency_i) + \beta_4 \ln(Monetary\ Value_i) + \varepsilon_i \quad (4a)$$

$$\ln(Average\ Sales\ Rank_i) = \alpha + \beta_1 Days\ Since\ First\ Internet\ Experience_i + \beta_2 \ln(Recency_i) + \beta_3 \ln(Frequency_i) + \beta_4 \ln(Monetary\ Value_i) + \varepsilon_i \quad (4b)$$

We first estimate the model using all consumers who purchased at least one item from the Internet channel but none from the catalog channel in 2002. Notice that "RFM" and *Days Since First Internet Purchase_i* measures are not defined for "new customers" who were acquired after January 1, 2002. After

¹⁰ *Recency_i* is defined as the number of days prior to January 1, 2002 that consumer *i* made a purchase. *Frequency_i* is defined as the number of items placed by the consumer prior to January 1, 2002. *Monetary Value_i* is defined as the average price of the items in consumer *i*'s historical orders.

omitting new customers, we have 66,975 consumers who purchased at least one item from the Internet channel but none from the catalog channel. In Table 3 we report estimates for Equations (4a) and (4b).

Table 3: The Effect of Consumers' Prior Internet Experience on Consumers' Tendency to Purchase Obscure Products through the Internet Channel

	Model (4a)	Model (4b)
Intercept	4.203** (0.057)	4.245** (0.055)
Recency	-0.001 (0.005)	-0.005 (0.004)
Frequency	0.213** (0.006)	0.209** (0.007)
Monetary Value	0.383** (0.010)	0.382** (0.010)
Prior Internet Experience	0.093** (0.015)	
Days Since First Internet Purchase		0.016** (0.002)
R ²	0.035	0.035
Sample Size	66,975	66,975

Standard errors are in parentheses.

**Significantly different from zero, $p < 0.01$.

*Significantly different from zero, $p < 0.05$.

The *Prior Internet Experience_i* coefficient is positive and highly significant, even after using consumers' historical purchasing characteristics as controls for consumer heterogeneity. It indicates that the average sales rank of products purchased by consumers with prior Internet experience is about 9.3% larger than that for consumers without prior Internet experience. Similarly, in the second model, the *Days Since First Internet Purchase_i* coefficient is also positive and highly significant. We interpret these results as evidence that Internet purchases made by consumers with prior Internet experience are skewed toward niche products. This empirical evidence is consistent with the prediction that purchases made by consumers with lower search costs are skewed towards niche products. We note that two control variables, *Frequency_i* and *Monetary Value_i* are also correlated with the dependent variable, indicating

Internet purchases made by consumers who have historically purchased more frequently and spent more money on this company's products are skewed toward niche products.

Ruling Out Consumer Heterogeneity as An Alternative Explanation

The “RFM” measures, used as control variables in our analysis, are also the standard control variables used by the industry and the academic literature to segment consumers and control for consumer heterogeneity. However, we would like to go one step further and test whether any consumer heterogeneity that are not controlled for by “RFM” measures could have caused the *Prior Internet Experience_i* coefficient to be positive and highly significant. If the positive coefficient on *Prior Internet Experience_i* merely reflects the effect of the uncontrolled-for consumer heterogeneity on the dependent variable *Average Sales Rank_i*, then we would expect that the effect of the uncontrolled-for consumer heterogeneity extends to the catalog channel as well. We would also expect that the positive coefficient on *Prior Internet Experience_i* extends to the catalog channel. Our unique data set allows use to test this, because we can study a common sample of consumers who have purchased at least one item from both the Internet channel and the catalog channel. We omit new customers acquired after January 1, 2002 for whom “RFM” measures are not defined. This leaves us with 37,836 consumers.

We first re-estimate the models in Equations (4a) and (4b) using purchases made by these “both-channel” consumers through the Internet channel. These results are reported in the first two columns of Table 4. They simply replicate the results reported in Table 3—the coefficients on *Prior Internet Experience_i* and *Days Since First Internet Purchase_i* are positive and highly significant.

We then repeat the analysis using the same sample of “both-channel” consumers but focusing on their purchases from the catalog channel. For each of the 37,836 consumers, we identify the products she has purchased through the catalog channel in 2002 and calculate the average sales rank of these products, where the sales rank for each product is calculated using aggregate sales in the catalog channel.¹¹ We then used this *Average Sales Rank_i* as the dependent variable and re-estimate the models in Equations (4a) and (4b). These results are reported in the last two columns of Table 4.

Table 4: The Effect of Consumers’ Prior Internet Experience on Consumers’ Tendency to Purchase Obscure Products through the Internet and Catalog Channels

	Internet		Catalog	
	Model (4a)	Model (4b)	Model (4a)	Model (4b)
Intercept	4.818** (0.074)	4.846** (0.073)	5.646** (0.053)	5.627** (0.052)
Recency	-0.010 (0.006)	-0.013* (0.005)	-0.023** (0.006)	-0.021** (0.004)
Frequency	0.136** (0.007)	0.133** (0.007)	0.101** (0.005)	0.102** (0.005)
Monetary Value	0.329** (0.013)	0.328** (0.013)	0.305** (0.009)	0.305** (0.009)
Prior Internet Experience	0.079** (0.017)		-0.044** (0.012)	
Days Since First Internet Purchase		0.014** (0.003)		-0.007** (0.002)
R ²	0.025	0.025	0.036	0.035
Sample Size	37,836	37,836	37,836	37,836

Standard errors are in parentheses.

**Significantly different from zero, $p < 0.01$.

*Significantly different from zero, $p < 0.05$.

In the last two columns of Table 4, the coefficient on *Prior Internet Experience_i* remains highly significant, but the coefficient is now negative. This indicates that, unlike their Internet purchases, the catalog purchases made by consumers with previous Internet experience are not skewed towards niche products. Similarly, the coefficient for the *Days Since First Internet Purchase_i* measure is also now

¹¹ By using sales ranks specific to each channel we control for differences in which products were featured as key products in each channel. Our results are robust to using a common (aggregate) sales ranking across the two channels.

negative and highly significant. These results help us rule out the explanation that the positive coefficient on *Prior Internet Experience_i* in the Internet channel merely reflects the effect of the uncontrolled-for consumer heterogeneity on the dependent variable.

In addition, the contrasting findings in the first two columns and last two columns in Tables 4 suggest that consumers who use both channels and have prior Internet experience tend to purchase more popular products through the catalog channel while they tend to purchase more niche products through the Internet channel, compared with consumers who have no prior Internet experience. The contrast in findings between the two channels is precisely the pattern we would expect if search costs are lower on the Internet (than through the catalog channel) and search costs are lower for consumers with prior Internet experience (than without prior Internet experience). When search costs are lower on the Internet than through the catalog channel, we would expect that consumers who use both channels will prefer to use the Internet channel when buying niche products. This selection-of-channel effect becomes larger for consumers with prior Internet experience than for consumers without prior Internet experience. We note that this selection-of-channel effect that happens to both-channel consumers is in fact a result of lower search costs on the Internet. But the findings shown in Table 4 are not consistent with the alternative explanation that differences in the concentration of product sales across channels and across individual consumers merely result from (controlled-for or uncontrolled-for) consumer heterogeneity.

5. Conclusions

Previous research has addressed several implications of lowering search costs in electronic markets: ranging from the impact on prices, price dispersion, and consumer price sensitivity. However, there has been little research studying the relationship between search costs and the concentration of product sales. This paper focuses on such a relationship and provides a demand-side explanation for why the niche products account for a larger percentage of sales on the Internet than in offline retail channels.

We first provide a theoretical model illustrating how a reduction in search costs may affect the concentration of products sales in the setting of a multi-product monopolist. We then present empirical

evidence that confirms the existence of the Long Tail on the Internet: the Internet channel exhibits a significantly less concentrated sales distribution when compared with the catalog channel, even though these two channels offer the same products at the same set of prices. One alternative explanation for this difference is that consumers who shop on the Internet are different from consumers who shop through the catalog channel. By focusing on a common sample of consumers purchase through both channels, we are able to control for this potential selection bias. We find that the difference in the concentration of sales across the two channels remains statistically significant.

We also exploit the variation in search costs across consumers by investigating whether measures of an individual consumer's search costs on the Internet correlate with her tendency to purchase niche products in each channel. We find evidence that Internet purchases made by consumers with prior Internet experience are more skewed toward niche products, compared with consumers who have no such experience – and this pattern is reversed in the catalog channel. This comparison addresses the concern that our previous results do not control for unobservable differences between the Internet and catalog channels in the process used to feature key products. It further validates our explanation that variations in search costs across consumers lead to variations in the concentration of product sales across consumers, by ruling out consumer heterogeneity as an alternative explanation.

As companies invest in ever-more sophisticated information technologies that allow consumers to actively and passively discover products that they otherwise would not have considered, and as consumers gain more experience using these IT-enabled tools, our findings suggest that product sales will become less and less concentrated. The balance of power will continue to shift from a few best-selling products to niche products that are previously difficult to be discovered by consumers. This Long Tail phenomenon will have a profound impact on a firm's product development strategy, operations strategy, and marketing strategy. Because the underlying technological drivers that we have studied in this paper are certain to continue to progress in advanced economies, the implications of these technologies for firm strategies and economic welfare are likely to become increasingly important.

References

- Anderson, Chris. 2004. "The Long Tail." *Wired Magazine*, October 2004.
- Anderson, Eric, and Duncan Simester. 2007. "Price Stickiness and Customer Antagonism," mimeo, MIT, Cambridge MA.
- Anderson, Simon and Regis Renault. 1999. Pricing, Product Variety, and Search Costs: A Bertrand-Chamberlin-Diamond Model. *RAND Journal of Economics* 30(4) 719-735.
- Bresnahan, Tim, and Robert J. Gordon ed., *The Economics of New Goods*, National Bureau of Economic Research Studies in Income and Wealth, Univ. of Chicago Press, 1997.
- Bakos, Yannis. 1999. Reducing Buyer Search Costs: Implications for Electronic Marketplaces. *Management Science* 43(12) 1676-1692.
- Brown, Jeffrey R. and Austan Goolsbee. 2002. Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry. *Journal of Political Economy* 110(3) 481-507.
- Brynjolfsson, Erik, Yu J. Hu, and Michael D. Smith. 2003 Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers. *Management Science* 49(11) 1580-1596.
- Brynjolfsson, Erik, Yu J. Hu, and Michael D. Smith. 2006. From Niches to Riches: The Anatomy of the Long Tail. *Sloan Management Review*. Summer 2006, 47(4). 67-71.
- Brynjolfsson, Erik, and Michael D. Smith. 2000. Frictionless Commerce? A Comparison of Internet and Conventional Retailers. *Management Science* 46(4) 563-585.
- Cachon, Gerard., Christian Terwiesch, and Yi Xu. 2006. On the Effects of Consumer Search and Firm Entry on Multiproduct Competition. Forthcoming. *Marketing Science*.
- Chevalier, Judith, and Austan Goolsbee. 2003. Measuring Prices and Price Competition Online: Amazon and Barnes and Noble. *Quantitative Marketing and Economics*. 1(2) 203-222.
- Clay, Karen, Ramayya Krishnan, Eric Wolff, and Danny Fernnades. 2003. Retail Strategies on the Web: Price and Non-Price Competition in the Online Book Industry. *Journal of Industrial Economics* 49(4) 521-540.
- Diamond, Peter. 1971. A Model of Price Adjustment. *Journal of Economic Theory* 3: 156-168.
- DVDStation. 2005. Studio Impact: Catalog and Long Tail Product Distribution at Retail. Available at <http://www.dvdeverywhere.com/blog/download/DVDStation-CatalogImpact.pdf>.
- Elberse, Anita and Felix Oberholzer-Gee (2007). Superstars and Underdogs: An Examination of The Long Tail Phenomenon in Video Sales. Working Paper 07-015, Harvard Business School.
- Garicano, Luis and Thomas N. Hubbard. 2006. The Return to Knowledge Hierarchies. Working paper, University of Chicago, Chicago IL.
- Ghose, Anindya and Bin Gu. 2006. Search Costs, Demand Structure and Long Tail in Electronic Markets: Theory and Evidence. Working Paper 06-19, NET Institute, New York University.
- Gini, Corrado. 1912. Variabilità e mutabilità. Reprinted in *Memorie di Metodologica Statistica* Editors: Pizetti, E. and T. Salvemini. Libreria Eredi Virgilio Veschi, Rome, Italy, 1955.

- Greco, Albert N. 1997. *The Book Publishing Industry*. Allyn and Bacon, Boston.
- Hann, Il-horn, Eric Clemons, and Lorin Hitt. 2003. Price Dispersion and Differentiation in Online Travel: An Empirical Investigation. *Management Science* 48(4) 534-549.
- Hoque, Abeer Y. and Gerald L. Lohse 1999. An Information Search Cost Perspective for Designing Interfaces for Electronic Commerce. *Journal of Marketing Research* 36(3) 387-394.
- Kohn, Meir G. and Steven Shavell. 1974. The Theory of Search. *Journal of Economic Theory* 9(2) 93-123.
- Lorenz, Max O. 1905. Methods of Measuring the Concentration of Wealth. *Publications of the American Statistical Association* 9(70) 209-219.
- Pareto, Vilfredo. 1896. Cours d'economie politique. Bousquet, G.H., G. Busino, eds. *Oevres Completes de Vilfredo Pareto*, 1. Librairie Droz, Geneva, 1964. Originally published 1896.
- Perloff, Jeremy, and Steven Salop. 1985. Equilibrium with Product Differentiation. *Review of Economic Studies* 12 107-120.
- Professional Publishing Report. 1999. University Presses Credit Internet for Increased Sales. Volume 3, Number 2, January 29.
- Schulz, Norbert, and Konrad Stahl. 1996. Do Consumers Search for the Highest Price? Oligopoly Equilibrium and Monopoly Optimum in Differentiated-products Market. *RAND Journal of Economics* 27(3) 542-562.
- Scott Morton, Fiona, Florian Zettelmeyer, and Jorge Silva-Risso. 2001. Internet Car Retailing. *Journal of Industrial Economics* 49(4) 501-520.
- Wolinsky, Asher. 1986. True Monopolistic Competition as A Result of Imperfect Information. *Quarterly Journal of Economics* 101(3) 493-512.
- Zipf, George. 1949. *Human Behavior and the Principle of Least Effort*. Addison-Wesley, Cambridge, MA.

Appendix

Proof of Claim 1.

First, Kohn and Shavell (1974) have proved that if one additional search yields a negative net benefit, then the net benefit from searching, even with the option of further searches after this one search, would be negative. Therefore, consumer i only focuses on one additional search and does not consider the value of further searches. Second, if consumer i has just found a niche product that fits her taste in the previous search, she will stop searching immediately. This is because finding a niche product that fits her taste is the best outcome she can hope for and her utility cannot be increased by searching further. Third, if consumer i has just found a niche product that does not fit her taste in the previous search, she will continue to search. This is because consumer i 's benefit from one additional search will remain unchanged, after finding a niche product that does not fit her taste. If it is optimal for her to conduct the previous search, it is optimal for her to conduct one additional search after the previous search.

Combining this result with the result that finding a niche product that fits consumer i 's taste, we have proved the following. If consumer i decides to conduct the first search of niche products, she will continue to search until she finds a niche product that fits her taste or all $n-k$ niche products are searched.

Case 1) None of the key products fits consumer i 's taste, i.e., $w_i = 0$.

In this case, consumer i 's expected benefit from the first search of niche products is

$(1-\alpha) * \max\{0, t_i - p_N\}$, while search cost is s . If $s \geq (1-\alpha)(1-p_N)$, search cost is higher than the expected benefit, for any $t_i \in [0,1]$. Thus, consumer i will not search niche products. She will exit the

market without purchasing. If $0 \leq s < (1-\alpha)(1-p_N)$, we know that $(1-\alpha) * \max\{0, t_i - p_N\} > s$ is

equivalent to $t_i > p_N + \frac{s}{1-\alpha}$. Thus, consumer i whose t_i satisfies $t_i > p_N + \frac{s}{1-\alpha}$ will search niche

products and purchase the first niche product that fits her taste. If no such product is found after searching

all the niche products, she will exit the market without purchasing. Consumer i whose t_i satisfies

$t_i \leq p_N + \frac{s}{1-\alpha}$ will not search niche products. She will exit the market without purchasing.

Case 2) One of the key products fits consumer i 's taste, i.e., $w_i = t_i$, and $p_N \geq p_K$.

In this case, consumer i 's expected benefit from the first search of niche products is

$(1-\alpha) * [\max\{0, t_i - p_N\} - \max\{0, t_i - p_K\}]$. This is non-positive because $p_N \geq p_K$. Thus, consumer i will not search niche products. Consumer i whose t_i satisfies $t_i > p_K$ will purchase the key product that gives her the highest value and exit the market. Consumer i whose t_i satisfies $t_i \leq p_K$ will exit the market without purchasing.

Case 3) One of the key products fits consumer i 's taste, i.e., $w_i = t_i$, and $p_N < p_K$.

In this case, consumer i 's expected benefit from the first search of niche products is

$(1-\alpha) * [\max\{0, t_i - p_N\} - \max\{0, t_i - p_K\}]$, which is equal to $(1-\alpha) * (p_K - p_N)$ for $t_i \in (p_K, 1]$,

$(1-\alpha) * (t_i - p_N)$ for $t_i \in [p_N, p_K]$, and 0 for $t_i \in [0, p_N]$. If $s \geq (1-\alpha)(p_K - p_N)$, search cost is higher than the expected benefit, for any $t_i \in [0, 1]$. Thus, consumer i will not search niche products. Consumer i whose t_i satisfies $t_i > p_K$ will purchase the key product that gives her the highest value and exit the market. Consumer i whose t_i satisfies $t_i \leq p_K$ will exit the market without purchasing. If

$0 \leq s < (1-\alpha)(p_K - p_N)$, we know that $(1-\alpha) * [\max\{0, t_i - p_N\} - \max\{0, t_i - p_K\}] > s$ is satisfied for

both $t_i \in (p_K, 1]$ and $t_i \in (p_N + \frac{s}{1-\alpha}, p_K]$. Thus, consumer i whose t_i satisfies $t_i > p_K$ will search niche

products and purchase the first niche product that fits her taste. If no such product is found after searching all the niche products, she will purchase the key product that gives her the highest value and exit the

market. Consumer i whose t_i satisfies $p_N + \frac{s}{1-\alpha} < t_i \leq p_K$ will search niche products and purchase the

first niche product that fits her taste. If no such product is found after searching all the niche products, she

will exit the market without purchasing. Consumer i whose t_i satisfies $t_i \leq p_N + \frac{s}{1-\alpha}$ will not search niche products. She will exit the market without purchasing. *Q.E.D.*

Proof of Claim 2.

We need to solve the firm's profit maximization problem for five cases.

Case A) $p_N < p_K$ and $s \geq (1-\alpha)(1-p_N)$. Consumers who have $w_i = 0$ will not search niche products, according to the result in Case 1) in Claim 1. Consumers who have $w_i = t_i$ will not search niche products, according to the result in Case 3) in Claim 1. The demand for each of the key products is

$D_K(p_K, p_N, s, n, k) = \frac{L}{k}(1-p_K)(1-\alpha^k)$, the demand for each of the niche products is

$D_N(p_K, p_N, s, n, k) = 0$, and the firm's profit function is $\pi(p_K, p_N, s, n, k) = L(p_K - c)(1-p_K)(1-\alpha^k)$.

Solving the firm's profit maximization problem— $\max_{p_K, p_N} \pi(p_K, p_N, s, n, k)$ s.t. $p_N < p_K$,

$s \geq (1-\alpha)(1-p_N)$, we have: $p_K^* = \frac{1+c}{2}$, $p_N^* \in [1 - \frac{s}{1-\alpha}, 1]$, $\pi^* = \frac{L(1-c)^2}{4}(1-\alpha^k)$, if

$s > \frac{(1-\alpha)(1-c)}{2}$; and $p_K^* = 1 - \frac{s}{1-\alpha} + \varepsilon$, $p_N^* \in [1 - \frac{s}{1-\alpha}, 1 - \frac{s}{1-\alpha} + \varepsilon)$,

$\pi^* = \frac{L}{4}(1 - \frac{s}{1-\alpha} + \varepsilon - c)(\frac{s}{1-\alpha} - \varepsilon)(1-\alpha^k)$, if $s \leq \frac{(1-\alpha)(1-c)}{2}$.

Case B) $p_N < p_K$ and $(1-\alpha)(p_K - p_N) \leq s < (1-\alpha)(1-p_N)$. Consumer i who have $w_i = 0$ and

$t_i > p_N + \frac{s}{1-\alpha}$ will search niche products, according to the result in Case 1) in Claim 1. Consumers who

have $w_i = t_i$ will not search niche products, according to the result in Case 3) in Claim 1. The demand for

each of the key products is $D_K(p_K, p_N, s, n, k) = \frac{L}{k}(1-p_K)(1-\alpha^k)$, the demand for each of the niche

products is $D_N(p_K, p_N, s, n, k) = \frac{L}{n-k}(1 - \frac{s}{1-\alpha} - p_N)\alpha^k(1-\alpha^{n-k})$, and the firm's profit function is

$\pi(p_K, p_N, s, n, k) = L(p_K - c)(1 - p_K)(1 - \alpha^k) + L(p_N - c)(1 - \frac{s}{1 - \alpha} - p_N)\alpha^k(1 - \alpha^{n-k})$. Solving the

firm's profit maximization problem— $\max_{p_K, p_N} \pi(p_K, p_N, s, n, k)$ s.t. $p_N < p_K$,

$(1 - \alpha)(p_K - p_N) \leq s < (1 - \alpha)(1 - p_N)$, we have: $p_K^* = \frac{1 + c}{2}$, $p_N^* = \frac{1}{2}(1 - \frac{s}{1 - \alpha} + c)$,

$\pi^* = \frac{L(1 - c)^2}{4}(1 - \alpha^k) + \frac{L}{4}(1 - \frac{s}{1 - \alpha} - c)^2\alpha^k(1 - \alpha^{n-k})$, if $s \leq (1 - \alpha)(1 - c)$. If $s > (1 - \alpha)(1 - c)$, there does

not exist p_K and p_N that satisfy $(1 - \alpha)(p_K - p_N) \leq s < (1 - \alpha)(1 - p_N)$.

Case C) $p_N < p_K$ and $s < (1 - \alpha)(p_K - p_N)$. Consumer i who have $w_i = 0$ and $t_i > p_N + \frac{s}{1 - \alpha}$ will

search niche products, according to the result in Case 1) in Claim 1. Consumers who have $w_i = t_i$ and

$t_i > p_N + \frac{s}{1 - \alpha}$ will search niche products, according to the result in Case 3) in Claim 1. The demand for

each of the key products is $D_K(p_K, p_N, s, n, k) = \frac{L}{k}(1 - p_K)(1 - \alpha^k)\alpha^{n-k}$, the demand for each of the

niche products is $D_N(p_K, p_N, s, n, k) = \frac{L}{n - k}(1 - \frac{s}{1 - \alpha} - p_N)(1 - \alpha^{n-k})$, and the firm's profit function is

$\pi(p_K, p_N, s, n, k) = L(p_K - c)(1 - p_K)(1 - \alpha^k)\alpha^{n-k} + L(p_N - c)(1 - \frac{s}{1 - \alpha} - p_N)(1 - \alpha^{n-k})$. Solving the

firm's profit maximization problem— $\max_{p_K, p_N} \pi(p_K, p_N, s, n, k)$ s.t. $p_N < p_K$, $s < (1 - \alpha)(p_K - p_N)$, we

have: $p_K^* = \frac{1}{2}(1 + c + \frac{s}{1 - \alpha} \frac{1 - \alpha^{n-k}}{1 - \alpha^n})$, $p_N^* = \frac{1}{2}(1 + c + \frac{s}{1 - \alpha} \frac{1 - \alpha^{n-k}}{1 - \alpha^n} - 2 \frac{s}{1 - \alpha})$,

$\pi^* = \frac{L}{4}[(1 - c)^2 - (\frac{s}{1 - \alpha} \frac{1 - \alpha^{n-k}}{1 - \alpha^n})^2](1 - \alpha^k)\alpha^{n-k}$, if $s \leq (1 - \alpha)(1 - c)$. If
 $+ \frac{L}{4}(1 - c + \frac{s}{1 - \alpha} \frac{1 - \alpha^{n-k}}{1 - \alpha^n} - 2 \frac{s}{1 - \alpha})(1 - c - \frac{s}{1 - \alpha} \frac{1 - \alpha^{n-k}}{1 - \alpha^n})(1 - \alpha^{n-k})$

$s > (1 - \alpha)(1 - c)$, there do not exist p_K and p_N that satisfy $s < (1 - \alpha)(p_K - p_N)$.

Case D) $p_N \geq p_K$ and $s > (1-\alpha)(1-p_N)$. Consumers who have $w_i = 0$ will not search niche products, according to the result in Case 1) in Claim 1. Consumers who have $w_i = t_i$ will not search niche products, according to the result in Case 2) in Claim 1. The demand for each of the key products is

$$D_K(p_K, p_N, s, n, k) = \frac{L}{k}(1-p_K)(1-\alpha^k), \text{ the demand for each of the niche products is}$$

$$D_N(p_K, p_N, s, n, k) = 0, \text{ and the firm's profit function is } \pi(p_K, p_N, s, n, k) = L(p_K - c)(1-p_K)(1-\alpha^k).$$

Solving the firm's profit maximization problem— $\max_{p_K, p_N} \pi(p_K, p_N, s, n, k)$ s.t. $p_N \geq p_K$,

$$s \geq (1-\alpha)(1-p_N), \text{ we have: } p_K^* = \frac{1+c}{2}, p_N^* \in \left[\frac{1+c}{2}, 1\right], \pi^* = \frac{L(1-c)^2}{4}(1-\alpha^k), \text{ if } s > \frac{(1-\alpha)(1-c)}{2};$$

$$\text{and } p_K^* = \frac{1+c}{2}, p_N^* \in \left(1 - \frac{s}{1-\alpha}, 1\right], \pi^* = \frac{L(1-c)^2}{4}(1-\alpha^k), \text{ if } s \leq \frac{(1-\alpha)(1-c)}{2}.$$

Case E) $p_N \geq p_K$ and $s \leq (1-\alpha)(1-p_N)$. Consumer i who have $w_i = 0$ and $t_i > p_N + \frac{s}{1-\alpha}$ will search niche products, according to the result in Case 1) in Claim 1. Consumers who have $w_i = t_i$ will not search niche products, according to the result in Case 2) in Claim 1. The demand for each of the key products is

$$D_K(p_K, p_N, s, n, k) = \frac{L}{k}(1-p_K)(1-\alpha^k), \text{ the demand for each of the niche products is}$$

$$D_N(p_K, p_N, s, n, k) = \frac{L}{n-k}\left(1 - \frac{s}{1-\alpha} - p_N\right)\alpha^k(1-\alpha^{n-k}), \text{ and the firm's profit function is}$$

$$\pi(p_K, p_N, s, n, k) = L(p_K - c)(1-p_K)(1-\alpha^k) + L(p_N - c)\left(1 - \frac{s}{1-\alpha} - p_N\right)\alpha^k(1-\alpha^{n-k}). \text{ Solving the}$$

firm's profit maximization problem— $\max_{p_K, p_N} \pi(p_K, p_N, s, n, k)$ s.t. $p_N \geq p_K, s < (1-\alpha)(1-p_N)$, we have:

$$p_K^* = \frac{1}{2}\left(1 - \frac{s}{1-\alpha} + c\right), p_N^* = \frac{1}{2}\left(1 - \frac{s}{1-\alpha} + c\right),$$

$$\pi^* = \frac{L}{4}[(1-c)^2 - (\frac{s}{1-\alpha})^2](1-\alpha^k) + \frac{L}{4}(1-\frac{s}{1-\alpha}-c)^2\alpha^k(1-\alpha^{n-k}), \text{ if } s < (1-\alpha)(1-c); \text{ and}$$

$$p_K^* = 1 - \frac{s}{1-\alpha}, p_N^* = 1 - \frac{s}{1-\alpha}, \pi^* = L(1-\frac{s}{1-\alpha}-c)\frac{s}{1-\alpha}(1-\alpha^k), \text{ if } s \geq (1-\alpha)(1-c).$$

If $s > (1-\alpha)(1-c)$, it is easy to verify that the strategy that maximizes the firm's profit is the strategies in

either Case A) or Case D). The firm's optimal strategy is to set $p_K^* = \frac{1+c}{2}$, $p_N^* \in [\frac{1+c}{2}, 1]$. The firm's

maximum profit function is $\pi^*(s, n, k) = \frac{L(1-c)^2}{4}(1-\alpha^k)$, the demand for each of the key products is

$$D_K^*(s, n, k) = \frac{L}{2k}(1-c)(1-\alpha^k), \text{ and the demand for each of niche products is } D_N^*(s, n, k) = 0.$$

If $s \leq (1-\alpha)(1-c)$, it is easy to verify that the strategy that maximizes the firm's profit is the strategy in

Case B). The firm's optimal strategy is to set $p_K^* = \frac{1+c}{2}$, $p_N^* = \frac{1}{2}(1-\frac{s}{1-\alpha}+c)$. The firm's maximum

profit function is $\pi^*(s, n, k) = \frac{L(1-c)^2}{4}(1-\alpha^k) + \frac{L}{4}(1-\frac{s}{1-\alpha}-c)^2\alpha^k(1-\alpha^{n-k})$, the demand for each of

the key products is $D_K^*(s, n, k) = \frac{L}{2k}(1-c)(1-\alpha^k)$, and the demand for each of niche products is

$$D_N^*(s, n, k) = \frac{L}{2(n-k)}(1-\frac{s}{1-\alpha}-c)(\alpha^k - \alpha^n). \text{ Q.E.D.}$$

Proof of Proposition 1.

The firm's maximum profit function is continuous at any $s \geq 0$ other than $s = (1-\alpha)(1-c)$. Because

$$\lim_{s \rightarrow (1-\alpha)(1-c)^-} \pi^*(s, n, k) = \lim_{s \rightarrow (1-\alpha)(1-c)^+} \pi^*(s, n, k) = \frac{L(1-c)^2}{4}(1-\alpha^k), \text{ it is also continuous at}$$

$$s = (1-\alpha)(1-c). \frac{\partial \pi^*(s, n, k)}{\partial s} = 0 \text{ for any } s > (1-\alpha)(1-c), \text{ and}$$

$$\frac{\partial \pi^*(s, n, k)}{\partial s} = \frac{L}{2} \left(1 - \frac{s}{1-\alpha} - c\right) \alpha^k (1 - \alpha^{n-k}) (-1) \frac{1}{1-\alpha} \leq 0 \text{ for any } s \leq (1-\alpha)(1-c). \text{ Therefore,}$$

$$\frac{\partial \pi^*(s, n, k)}{\partial s} \leq 0 \text{ for any } s \geq 0. \text{ Q.E.D.}$$

Proof of Proposition 2.

If $s > (1-\alpha)(1-c)$, the demand for each of the key products is $D_K^*(s, n, k) = \frac{L}{2k} (1-c)(1-\alpha^k)$, and the

demand for each of niche products is $D_N^*(s, n, k) = 0$. It is obvious that $D_K^*(s, n, k) \geq D_N^*(s, n, k)$.

Thus, the Gini Coefficient is

$$G(s, n, k) = 1 - \frac{2 \sum_{j=n-k+1}^n [(n+1-j) D_K^*(s, n, k)]}{(n+1)k D_K^*(s, n, k)} = 1 - \frac{k+1}{n+1}.$$

If $s \leq (1-\alpha)(1-c)$, the demand for each of the key products is $D_K^*(s, n, k) = \frac{L}{2k} (1-c)(1-\alpha^k)$, and the

demand for each of niche products is $D_N^*(s, n, k) = \frac{L}{2(n-k)} \left(1 - \frac{s}{1-\alpha} - c\right) (\alpha^k - \alpha^n)$. We first prove that

the demand for each of the featured products is higher than the demand for each of the niche products, i.e.,

$$D_K^*(s, n, k) \geq D_N^*(s, n, k).$$

Note that function $Y_1(x) = -x \ln(x) - 1 + x, x \in [0,1]$ reaches its maximum at $x=1$, because

$$Y_1'(x) = -\ln(x) \geq 0. \text{ Since } Y_1(1) = 0, \text{ we have } Y_1(x) \leq 0. \text{ Function } Y_2(l) = \frac{1-\alpha^l}{l}, \alpha \in [0,1], l > 0 \text{ is non-}$$

increasing with l , because its first-order derivative $Y_2'(l) = \frac{-\alpha^l \ln[\alpha^l] - 1 + \alpha^l}{l^2} \leq 0$. Therefore, we have

$$\frac{1-\alpha^k}{k} \geq \frac{1-\alpha^n}{n}, \text{ which is equivalent to } \frac{1-\alpha^k}{k} \geq \frac{\alpha^k - \alpha^n}{n-k}. \text{ In addition, we have } 1-c \geq 1 - \frac{s}{1-\alpha} - c.$$

Thus, we have proved $D_K^*(s, n, k) \geq D_N^*(s, n, k)$.

The Gini Coefficient is

$$G(s, n, k) = 1 - \frac{2 \sum_{j=1}^{n-k} [(n+1-j)D_K^*(s, n, k)] + 2 \sum_{j=n-k+1}^n [(n+1-j)D_K^*(s, n, k)]}{(n+1)[(n-k)D_N^*(s, n, k) + kD_K^*(s, n, k)]}$$

$$= 1 - \frac{k+1}{n+1} - \frac{n(1 - \frac{s}{1-\alpha} - c)(\alpha^k - \alpha^n)}{(n+1)[(1 - \frac{s}{1-\alpha} - c)(\alpha^k - \alpha^n) + (1-c)(1-\alpha^k)]}$$

The Gini Coefficient function is continuous at any $s \geq 0$ other than $s = (1-\alpha)(1-c)$. Because

$$\lim_{s \rightarrow (1-\alpha)(1-c)^-} G(s, n, k) = \lim_{s \rightarrow (1-\alpha)(1-c)^+} G(s, n, k) = 1 - \frac{k+1}{n+1}, \text{ it is also continuous at } s = (1-\alpha)(1-c). \text{ Since}$$

$$\frac{\partial G(s, n, k)}{\partial s} = 0 \text{ for any } s > (1-\alpha)(1-c), \text{ and}$$

$$\frac{\partial G(s, n, k)}{\partial s} = \frac{n}{n+1} \frac{\frac{1}{1-\alpha}(\alpha^k - \alpha^n)(1-c)(1-\alpha^k)}{[(1 - \frac{s}{1-\alpha} - c)(\alpha^k - \alpha^n) + (1-c)(1-\alpha^k)]^2} \geq 0 \text{ for any } s \leq (1-\alpha)(1-c), \text{ we have}$$

$$\frac{\partial G(s, n, k)}{\partial s} \geq 0 \text{ for any } s \geq 0. \text{ Q.E.D.}$$

Proof of Proposition 3.

We will first calculate the demand for key products and niche products under different $s_i (i=1,2)$, p_K and p_N . The demand functions have already been calculated in the proof of Claim 2. To summarize, we have:

	Key Product Demand $D_K(p_K, p_N, s_i, n, k)$	Niche Product Demand $D_N(p_K, p_N, s_i, n, k)$
1) $p_N \geq p_K$		
1a) $s_i \geq (1-\alpha)(1-p_N)$	$\frac{L}{k}(1-p_K)(1-\alpha^k)$	0
1b) $s_i < (1-\alpha)(1-p_N)$	$\frac{L}{k}(1-p_K)(1-\alpha^k)$	$\frac{L}{n-k}(1 - \frac{s_i}{1-\alpha} - p_N)\alpha^k(1-\alpha^{n-k})$

2) $p_N < p_K$

$$\mathbf{2a)} \quad s_i \geq (1-\alpha)(1-p_N) \quad \frac{L}{k}(1-p_K)(1-\alpha^k) \quad 0$$

$$\mathbf{2b)} \quad \begin{array}{l} (1-\alpha)(p_K - p_N) \leq s_i \\ < (1-\alpha)(1-p_N) \end{array} \quad \frac{L}{k}(1-p_K)(1-\alpha^k) \quad \frac{L}{n-k}\left(1-\frac{s_i}{1-\alpha}-p_N\right)\alpha^k(1-\alpha^{n-k})$$

$$\mathbf{2c)} \quad s_i < (1-\alpha)(p_K - p_N) \quad \frac{L}{k}(1-p_K)(1-\alpha^k)\alpha^{n-k} \quad \frac{L}{n-k}\left(1-\frac{s_i}{1-\alpha}-p_N\right)(1-\alpha^{n-k})$$

Case 1) $p_N \geq p_K$

First, $\frac{\partial D_N(p_K, p_N, s_i, n, k)}{\partial s} = 0$ for any $s_i \geq (1-\alpha)(1-p_N)$; and

$$\frac{\partial D_N(p_K, p_N, s_i, n, k)}{\partial s} = \frac{L}{n-k}\alpha^k(1-\alpha^{n-k})\frac{-1}{1-\alpha} < 0 \text{ for any } s_i < (1-\alpha)(1-p_N). \text{ Therefore,}$$

$$\frac{\partial D_N(p_K, p_N, s_i, n, k)}{\partial s} \leq 0 \text{ for any } s \geq 0.$$

$$\text{Next, } G(p_k, p_N, s_i, n, k) = 1 - \frac{2 \sum_{j=n-k+1}^n [(n+1-j)D_K(p_k, p_N, s_i, n, k)]}{(n+1)kD_K(p_k, p_N, s_i, n, k)} = 1 - \frac{k+1}{n+1} \text{ when}$$

$$s_i \geq (1-\alpha)(1-p_N). \text{ Thus, } \frac{\partial G(p_K, p_N, s_i, n, k)}{\partial s} = 0 \text{ when } s_i \geq (1-\alpha)(1-p_N).$$

$$\begin{aligned} G(p_k, p_N, s_i, n, k) &= 1 - \frac{2 \sum_{j=1}^{n-k} [(n+1-j)D_N(p_k, p_N, s_i, n, k)] + 2 \sum_{j=n-k+1}^n [(n+1-j)D_K(p_k, p_N, s_i, n, k)]}{(n+1)[(n-k)D_N(p_k, p_N, s_i, n, k) + kD_K(p_k, p_N, s_i, n, k)]} \\ &= 1 - \frac{k+1}{n+1} - \frac{n(n-k)D_N(p_k, p_N, s_i, n, k)}{(n+1)[(n-k)D_N(p_k, p_N, s_i, n, k) + kD_K(p_k, p_N, s_i, n, k)]} \end{aligned}$$

when $s_i < (1-\alpha)(1-p_N)$. Since $\frac{\partial D_N(p_K, p_N, s_i, n, k)}{\partial s} < 0$ and $\frac{\partial D_K(p_K, p_N, s_i, n, k)}{\partial s} = 0$, we have

$$\frac{\partial G(p_K, p_N, s_i, n, k)}{\partial s} > 0 \text{ when } s_i < (1-\alpha)(1-p_N). \text{ Therefore, } \frac{\partial G(p_K, p_N, s_i, n, k)}{\partial s} \geq 0 \text{ for any } s \geq 0.$$

Case 2) $p_N < p_K$

First, $\frac{\partial D_N(p_K, p_N, s_i, n, k)}{\partial s} = 0$ for any $s_i \geq (1-\alpha)(1-p_N)$;

$$\frac{\partial D_N(p_K, p_N, s_i, n, k)}{\partial s} = \frac{L}{n-k} \alpha^k (1-\alpha^{n-k}) \frac{-1}{1-\alpha} < 0 \text{ for any } (1-\alpha)(p_K - p_N) \leq s_i < (1-\alpha)(1-p_N);$$

$$\frac{\partial D_N(p_K, p_N, s_i, n, k)}{\partial s} = \frac{L}{n-k} (1-\alpha^{n-k}) \frac{-1}{1-\alpha} < 0 \text{ for any } s_i < (1-\alpha)(p_K - p_N). \text{ Therefore,}$$

$$\frac{\partial D_N(p_K, p_N, s_i, n, k)}{\partial s} \leq 0 \text{ for any } s \geq 0.$$

$$\text{Next, } G(p_K, p_N, s_i, n, k) = 1 - \frac{2 \sum_{j=n-k+1}^n [(n+1-j)D_K(p_K, p_N, s_i, n, k)]}{(n+1)kD_K(p_K, p_N, s_i, n, k)} = 1 - \frac{k+1}{n+1} \text{ when}$$

$$s_i \geq (1-\alpha)(1-p_N). \text{ Thus, } \frac{\partial G(p_K, p_N, s_i, n, k)}{\partial s} = 0 \text{ when } s_i \geq (1-\alpha)(1-p_N).$$

$$\begin{aligned} G(p_K, p_N, s_i, n, k) &= 1 - \frac{2 \sum_{j=1}^{n-k} [(n+1-j)D_N(p_K, p_N, s_i, n, k)] + 2 \sum_{j=n-k+1}^n [(n+1-j)D_K(p_K, p_N, s_i, n, k)]}{(n+1)[(n-k)D_N(p_K, p_N, s_i, n, k) + kD_K(p_K, p_N, s_i, n, k)]} \\ &= 1 - \frac{k+1}{n+1} - \frac{n(n-k)D_N(p_K, p_N, s_i, n, k)}{(n+1)[(n-k)D_N(p_K, p_N, s_i, n, k) + kD_K(p_K, p_N, s_i, n, k)]} \end{aligned}$$

when $(1-\alpha)(p_K - p_N) \leq s_i < (1-\alpha)(1-p_N)$ or when $s_i < (1-\alpha)(p_K - p_N)$. Since

$$\frac{\partial D_N(p_K, p_N, s_i, n, k)}{\partial s} < 0 \text{ and } \frac{\partial D_K(p_K, p_N, s_i, n, k)}{\partial s} = 0, \text{ we have } \frac{\partial G(p_K, p_N, s_i, n, k)}{\partial s} > 0 \text{ when}$$

$(1-\alpha)(p_K - p_N) \leq s_i < (1-\alpha)(1-p_N)$ or when $s_i < (1-\alpha)(p_K - p_N)$. Therefore,

$$\frac{\partial G(p_K, p_N, s_i, n, k)}{\partial s} \geq 0 \text{ for any } s \geq 0. \text{ Q.E.D.}$$