

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

This version: December 2006

First version: December 2005

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

ABSTRACT

Product variety is an important component of consumer welfare, yet 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 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 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.

Acknowledgements: We thank Chris Anderson, Jerry Hausman, Lorin Hitt, seminar participants at Purdue University, 2005 Workshop on Information Systems and Economics (WISE) 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>

** Krannert School of Management, Purdue University; email: yuhu@purdue.edu

*** MIT Sloan School of Management; email: simester@mit.edu

1. Introduction

Much of the growth in consumer welfare in recent years is attributable to increases in product variety and the availability of new goods and services (Bresnahan and Gordon 1997). For example, while the Internet has lowered prices significantly in the book market, the welfare benefits from consumers' increased ability to search for and find a broader variety of titles is 5-7 times more important than the lower prices (Brynjolfsson, Hu and Smith 2003). As various information technologies continue to lower search costs, product variety may increase and the distribution of product sales across different products may change accordingly.

Historically, many markets have traditionally been dominated by a few best-selling and featured products. For example, book sales are concentrated on a few books by several established best-selling authors (Greco 1997). Similarly, Billboard "top 40" hits account for the lion's share on many over-the-air radio playlists and CD sales, and movie rentals are dominated by a few "new releases". Economists and business managers often use the Pareto Principle, also known as the 80/20 rule, to describe this phenomenon of sales concentration. This rule says a small proportion (e.g., 20 percent) of products in a market often generate a large proportion (e.g., 80 percent) of sales in the market.¹ However, the Internet has the potential to shift this balance. By greatly lowering search costs, information technology in general and Internet markets in particular may radically increase the collective share of niche products, and flatten the sales distribution. Anderson (2004) has coined a term—"The Long Tail"—to describe the phenomenon that niche products can make up a significant share of total sales. It is possible that the Pareto Principle has given way to the "Long Tail" on the Internet.

Anecdotal evidence suggests that Internet markets has helped shift the balance of power from featured products to niche products that are previously obscure. For example, Frank Urbanowski, Director of MIT Press, observed that the increased accessibility to backlist titles through the Internet has resulted in a 12%

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

increase in sales of these titles (Professional Publishing Report 1999). This increase has happened despite flat growth in overall book sales. Brynjolfsson, Hu and Smith (2003) find that obscure book titles, which are not even stocked by typical conventional bookstores because of their low sales, accounted for about 40% of Amazon.com's book sales revenue in year 2000. Similar observations have been made in electronic markets for music, DVDs, and electronics. For example, Rhapsody, an online music provider, streams more songs each month beyond its top 10,000 than it does its top 10,000. While newly released movies account for a dominant share of revenue in a video rental shop, DVDStation, a company that allows consumers to 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).

Understanding how Internet markets have affected the concentration of product sales has important managerial implications for a single multi-product firm. By lowering buyer search costs, Internet markets may expand the set of products buyers consider when making their purchases. Buyers are no longer limited to featured products that are heavily promoted, advertised and, as a result, highly visible. The expansion of buyers' choice sets leads to changes in the firm's profit function and changes in the firm's strategies in producing and marketing its products. As buyer search costs are lowered, the firm will benefit more from producing niche products that previously sold very little and there is a larger incentive to invest in the development of new products and new varieties. At the same time, the firm has a relatively smaller incentive to invest in the marketing of featured products. These long-term changes in firm strategies can lead to further changes in sales concentration (Brynjolfsson, Hu and Smith 2006).

Remarkably, despite the abundance of anecdotal evidence and the important economic consequences, theoretical and empirical research on how lower search costs on the Internet has changed the distribution of product sales and market structure 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, 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 these models do not predict a change in sales concentration. Our paper considers a model that allows heterogeneity in search costs across products (and channels), by assuming that search cost is low for one category of products and high for another category. A reduction in search costs affects these products asymmetrically, leading to a change in the distribution of product sales.

There is a growing body of empirical research that studies 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, there have been no investigations of how lower search costs affect the concentration of product sales. In this paper we explicitly model consumer search behavior to investigate how search costs affect the distribution of sales. We propose use of the Gini Coefficient and Pareto Curve, concepts that are traditionally used to study the inequality in income and wealth distribution (Lorenz 1905, Gini 1912, Pareto 1896), to measure sales concentration.

In the empirical section of this paper we test our theoretical predictions by comparing product sales through the Internet channel with sales through a non-Internet (catalog) channel. The two channels use the same order fulfillment methods and facilities, which controls for differences in sales tax policies, shipping costs, product selection, and the possibility of stockouts. Nonetheless, these two channels exhibit a significant difference in sales concentration, even after we control for consumer selection bias arising from differences in consumers who shop through the Internet and non-Internet channels. In addition, we use an individual-level analysis to show that consumers who have more experience using the Internet channel (and presumably have lower search costs) are more likely to purchase niche and obscure products, compared with consumers who have little or no experience in using the Internet channel. We observe the opposite outcome when comparing purchases by the same consumers through the catalog channel.

In Section 2 we begin with a model to illustrate how a change in search costs is expected to affect the distribution of product sales. This model also allows us to discuss how a reduction in search costs can

decrease a multi-product firm's incentives to invest in the marketing of featured products and increase its incentives to expand its product line. In Section 3, we present empirical evidence that lower search costs on the Internet have reduced the concentration in product sales. The paper concludes in Section 4 with a review of the findings and implications.

2. Model

We consider a multi-product monopoly firm that sells n ($n \gg 1$) products to consumers. Among the firm's n products, $k \in (1, n)$ of them are "featured" products, meaning products that are advertised by the firm and/or products that are displayed at featured locations. Examples 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 featured products. In contrast, consumers have to exert effort and incur search costs in order to find the rest of the products; we will call these $n-k$ products "niche" products. Specifically, we assume featured products have a zero search cost, and niche products have a positive search cost s .

The number of available products (n) and the number of featured products (k) depend on the firm's product development and marketing strategies. Treating these strategies as quasi-fixed, we will first solve for a short-term equilibrium in which the firm holds its product development strategies and marketing strategies constant, even when search costs change. That is, the number of available products (n) and the number of featured products (k) are held constant. In this equilibrium, the firm only considers its variable cost of production, $c \in (0, 1)$, which is assumed to be constant for all n products. Later we will solve for a long-term equilibrium in which the firm chooses the optimal n and k such that the firm's profit is maximized. This long-term equilibrium addresses how changes in consumer search costs affect the firm's product development and marketing strategies. Let the firm's cost of developing n products and making them available be $R(n)$, where $R' > 0$, $R'' \geq 0$, and the firm's cost of marketing k products and converting them from niche products to featured products be $M(k)$, where $M' > 0$, $M'' \geq 0$.

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.

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 learns about a different product and the value she places on the product, until all n products are searched. She can proceed to purchase any one of the products she has already searched without incurring any additional cost.² Each time a consumer searches, she finds a product that does not fit her taste with probability $\alpha \in (0,1)$ and finds a product that fits her taste with probability $1 - \alpha$. Consumer i places zero value ($v_{ij} = 0$) on product j if product j does not fit her taste, and she places value t_i if it does ($v_{ij} = t_i$). 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; after purchasing one unit of any of the n products, she exits the market.

Note that the firm will prefer to charge the same price for all featured products, and we use p_F 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 featured products (k), the search cost for niche products

² This specification of search without replacement and with recall is consistent with the previous literature on this topic. 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.

(s), the probability she will find a product that does not fit her taste every time she searches (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}); this is revealed only after consumer i has searched and evaluated product j .

Consumers' Search and Purchasing Strategies

Because featured products have zero search costs, a consumer will always search through all of the k featured 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 featured product with the highest value. Let w_i be the value that consumer i places on this product. Because w_i is 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 making a purchase; exiting the market after purchasing the featured product that gives her the highest value amongst all featured products; or continuing to search in the hope of finding a product that gives her higher utility (either a lower price or a higher value) than the most desirable product she has located so far.

Because the benefit from searching niche products is a function of w_i , α , t_i , p_F , and p_N , a consumer's decision to search a niche product depends upon these variables. The following claim describes the search and purchase strategies that maximize expected utility.

Claim 1

1. When none of the featured 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 $0 \leq s < (1 - \alpha)(1 - p_N)$, consumer i searches the niche product iff

$t_i > p_N + \frac{s}{1-\alpha}$ and purchases 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 making a purchase.

2. One of the featured products fits consumer i 's taste ($w_i = t_i$) and $p_N \geq p_F$

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

3. One of the featured products fits consumer i 's taste ($w_i = t_i$) and $p_N < p_F$

There are four scenarios to consider here. If $s \geq (1-\alpha)(p_F - p_N)$, consumer i will not search niche products. As long as $t_i > p_F$ she will purchase the featured product that gives her the highest value; otherwise she will exit the market without purchasing. If $0 \leq s < (1-\alpha)(p_F - p_N)$ and $t_i > p_F$

consumer i searches and purchases the first niche product that fits her taste. If no such niche product is found, she will purchase the featured product that offers the highest value. If $p_N + \frac{s}{1-\alpha} < t_i \leq p_F$

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. Finally, if $t_i \leq p_N + \frac{s}{1-\alpha}$ consumer i will not search niche products and instead simply exits the market without a purchase.

A complete proof of this claim, and the claims and propositions that follow, 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.

Short-term Equilibrium

We first solve for a short-term equilibrium in which the firm treats the number of available products (n) and number of featured products (k) as given parameters. The firm chooses the price for featured products (p_F) and the price for niche products (p_N) such that the firm's profit is maximized. Claim 2 characterizes the demand for the firm's products and the firm's profit under such an equilibrium.

Claim 2

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

	High Search Costs $s > (1-\alpha)(1-c)$	Low Search Costs $s \leq (1-\alpha)(1-c)$
Price of Featured Products P_F^*	$\frac{1+c}{2}$	$\frac{1+c}{2}$
Price of Niche Products P_N^*	$[0,1]$	$\frac{1}{2}\left(1-\frac{s}{1-\alpha}+c\right)$
Featured 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 if customers find a match amongst the featured products they do not search the niche products. Therefore, expected demand for the featured products is unaffected by the cost of searching niche products. However, as we would expect, expected demand for the niche products is strictly higher when search costs are lower. As a result, expected profits are also higher when search costs are lower (notice that $\alpha^k > \alpha^n \forall k > n$). We restate this result as a formal proposition.

Proposition 1. The firm's equilibrium profit is non-increasing with consumer search costs:

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

Next we consider the effect of search costs on the distribution of product sales. Economists have long used the Lorenz Curve and Gini Coefficient to describe the inequality in income and 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 or 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. The Gini Coefficient measuring sales concentration is non-decreasing with consumer

search costs: $\frac{\partial G(s, n, k)}{\partial s} \geq 0$ for any $s \geq 0$.

Sketch of Proof.

If search costs are high, $s > (1 - \alpha)(1 - c)$, the Gini Coefficient is:

$$G(s, n, k) = 1 - \frac{2 \sum_{j=n-k+1}^n [(n+1-j)D_F^*(s, n, k)]}{(n+1)kD_F^*(s, n, k)} = 1 - \frac{k+1}{n+1}. \quad (3)$$

If search costs are low, $s \leq (1 - \alpha)(1 - c)$, the Gini Coefficient is:

$$\begin{aligned}
G(s, n, k) &= 1 - \frac{2 \sum_{j=1}^{n-k} [(n+1-j)D_N^*(s, n, k)] + 2 \sum_{j=n-k+1}^n [(n+1-j)D_F^*(s, n, k)]}{(n+1)[(n-k)D_N^*(s, n, k) + kD_F^*(s, n, k)]} \\
&= 1 - \frac{k+1}{n+1} \frac{n(1 - \frac{s}{1-\alpha} - c)(\alpha^k - \alpha^n)}{(n+1)[(1 - \frac{s}{1-\alpha} - c)(\alpha^k - \alpha^n) + (1-c)(1 - \alpha^k)]}
\end{aligned} \tag{4}$$

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

As search costs decrease, 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. We note that this result is based on a short-term equilibrium, in which the firm does not adapt to lower search costs by changing its product development and marketing strategies.

The Effect of Search Costs in a Long-term Equilibrium

In the long-term, the firm may change its marketing and product development strategies.³ Specifically, the firm can choose n and k such that the firm's profit is maximized. The firm's long term profit maximization problem becomes:

$$\pi^*(s) = \max_{p_N, p_F, n, k} [\pi^*(s, n, k, p_N, p_F) - R(n) - M(k)]. \tag{5}$$

Proposition 3. The optimal number of available products that the firm offers in the long-term equilibrium

is non-increasing with consumer search costs: $\frac{\partial n^*(s)}{\partial s} \leq 0$ for any $s \geq 0$.

In the long-term equilibrium, when marketing and development strategies are not fixed, the firm will adapt to lower search costs by increasing the number of available products. As search costs fall, consumers are more likely to search, which leads to a larger incentive for the firm to invest in the development of new products and varieties.

³ It is possible that there are other constraints which prevent the firm from changing these strategies even in the long run. For instance, a multi-channel retailer may find it optimal to maintain a consistent marketing and development strategy across all channels, even if search costs are very different between channels.

Proposition 4. The optimal number of featured products that the firm offers in the long-term equilibrium

is non-decreasing with consumer search costs: $\frac{\partial k^*(s)}{\partial s} \geq 0$ for any $s \geq 0$.

The firm will also decrease the number of featured products. When search costs are high, consumers tend not to search. As a result, the firm's investment in marketing featured products can effectively help consumers discover products and increase the firm's demand. But as search costs fall, consumers are more likely to search through the broader set of products offered by the firm. As a result, the firm has less incentive to investment in the marketing of featured products.

Propositions 3 and 4 provide clear predictions about how the number of featured and niche products will change in a long-run equilibrium as search costs decrease: the number of niche products increases, while the number of featured products falls. The net impact on sales concentration is less clear. Notice first that if search costs are high, so that no customers search the niche products, the change in the cost of searching niche products initially does not affect the number of featured or niche products. The level of sales concentration is also unaffected.

However, when search costs are low, so that at least some customers purchase niche products, the outcome is ambiguous. By reducing the number of featured products, the reduction in search costs tends to increase the sales concentration amongst the remaining featured products. However, the aggregate sales of the diminishing pool of featured products also falls, as some of the demand is shifted to niche products. This yields two opposing effects on the sales concentration. If we calculated the Gini coefficient for just those featured products that remain, the coefficient would increase as search costs fall. However, if we consider both featured and niche products the net outcome depends on the relative magnitude of the two affects.

Similarly, the increased willingness of customers to search niche products when the cost of searching is lower increases the aggregate sales of niche products. However, there is also an increase in the number of

niche products, and so these additional sales are spread over more products. The net outcome is again unclear.

Summary

In the short-run equilibrium, an increase in search costs unambiguously leads to an increase in the dispersion of sales across products, diminishing the level of sales concentration. However, in the long-run firms may react to the change in search costs by increasing the proliferation of products and reducing their investments in featuring individual products. As a result, we cannot make a clear prediction about the long-run relationship between search costs and the distribution of product sales.

In the next section we use transaction data from a retailer to compare the distribution of sales across Internet and catalog channels. Because the two channels share the same product range and fulfillment facilities, the comparison controls for differences in the size of the product range and ensures that the outcome is not confounded by logistical considerations. We begin the next section by describing the company and discussing whether the difference between the sales distributions in the two channels should be interpreted as a test of the short or long-run equilibrium results.

3. Empirical Analyses

The company is a medium-sized retailing company that sells mainly women's clothing in the moderate price range.⁴ All of the products carry the company's private label brand and are sold exclusively through the company's catalog channel (mail and telephone orders), Internet website and a small number of retail stores. We have the transaction data of this company's consumers from year 1998, when the firm first started selling over the Internet, through year 2002. We do not have a record of purchases made by consumers in the company's retail stores, because the company was unable to adequately identify consumers purchasing in its stores. This limits our study to the Internet and catalog channels.

Linking Theoretical Model to Empirical Analyses

⁴ The company asked to remain anonymous.

A key feature of the company is that it maintains a policy of offering exactly the same product selection (and prices) through its Internet and catalog channels. This policy simplifies the firm'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 2006). Although there may be differences in search costs across the two channels, the firm does not respond by changing the number of available products.

In addition to offering the same products through both channels, the firm 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.

There are differences across the channels in the process used to feature products, together with the cost of featuring additional products. In the catalog channel 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. The firm controls which products appear in each catalog, the location of the products, the number of pages in each catalog, along with how many consumers receive each catalog. Featuring products on the Internet channel also involves product location. The firm decides which products appear on the "Home" page and the first page of each successive category and subcategory. It is difficult to compare the relative effectiveness of these "featuring" mechanisms across the two channels. Moreover, while the cost of these mechanisms vary, evaluating the relative magnitude of these costs is also difficult.

Recognizing that the number of featured products varies across the channels, together with the definition of what it means to be "featured", has important implications for the interpretation of our empirical results. When we compare the sales distributions across the Internet and catalog channels, we cannot distinguish whether differences in the sales distribution are due to differences in the number of featured products in

each channel or differences in the cost of search. 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. In practice, the evidence elsewhere that search costs on the Internet are lower than in a catalog (see for example Hoque and Lohse 1999) in part results from firms' efforts to facilitate customer search on their websites. If a website has more effective search and recommendation tools, search costs will be lower, while omitting these features will tend to increase search costs.

While the distinction between search costs and the number of featured products may be ambiguous, the link to our theoretical model is otherwise clear. We should think of this as a long-run equilibrium in which the number of products is fixed, but consumers' search costs and the number of featured products may vary. It remains an empirical question as to whether this leads to an increase or decrease in the concentration of product sales.

We will present results that include analysis of how the sales distribution varies across consumers with different search costs. These individual-level results help to isolate the effect of search costs from the effect of the number of featured products, because they focus on purchases made through only one channel. The results can be interpreted as providing further evidence about whether lower search costs lead to less concentration in product sales.

Initial Results

While we have data from 1998 through 2002, our analyses focus on transactions that occurred in 2002, because there were relatively few Internet transactions in earlier years. We will later use the earlier data to control for customer differences.

Table 1 summarizes the comparison of the Internet channel and catalog channel, in terms of the number of items sold and the number of unique consumers who made purchases in 2002. We see that during 2002 the company sold over ten times as many items through the catalog channel as through the Internet channel. There were also more individual consumers who made at least one purchase through the catalog channel.

Table 1: Size of Internet Channel and Catalog Channel in 2002

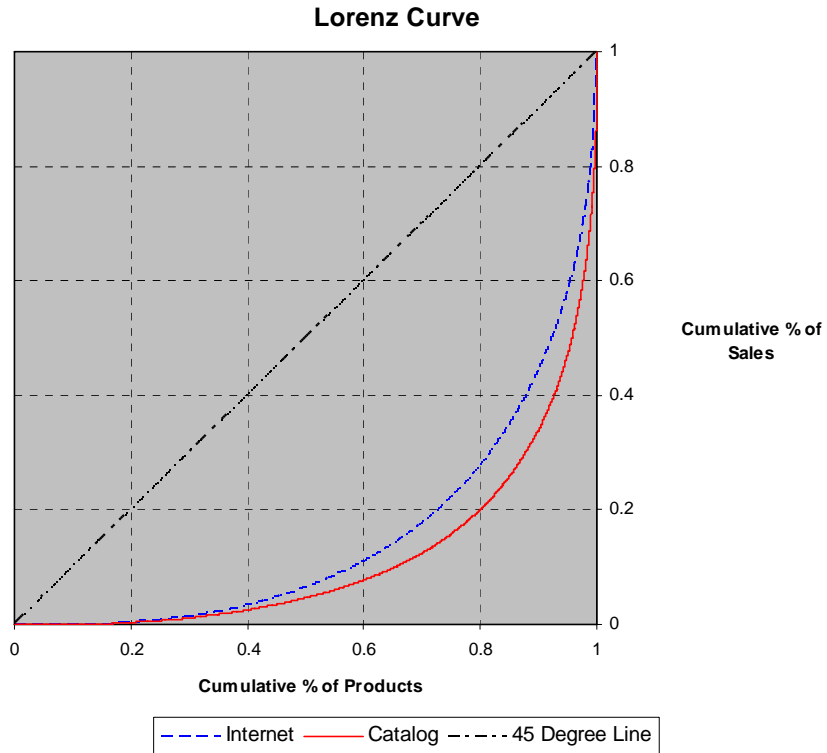
	Internet	Catalog
Items Sold	702,659	7,725,574
Unique Consumers Who Purchased	220,028	762,499

In comparing the distribution of sales we want to ensure that the findings are not affected by the differences in aggregate demand. Therefore, we randomly selected 100,000 transactions from each channel to represent demand in that channel (where each transaction represents the purchase of a single item). Overall, these 100,000 transactions included 5,568 different 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 constructed Lorenz Curves and calculated Gini Coefficients for both channels using all 6,550 products. Reassuringly, the findings that follow are all robust to either using all of the transactions in the two samples, 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) 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 more equal 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). From the Lorenz Curve, we can easily obtain the percentage of total sales generated by the bottom 80% products. For the catalog channel, the top 20% of products generate

79.9% of sales; corresponding with remarkable precision to the 80/20 rule. For the Internet channel, the top 20% of products generate 72.3% of sales.

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 Curve 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 such a difference is statistically significant. In order to do so, we will fit the sales and sales rank data to a log-linear relationship and compare the coefficient on sales rank that is obtained using Internet channel data with the coefficient on sales rank that is obtained using catalog channel data. In particular we estimated the following log-linear relationship:

$$\ln(\text{Sales}_j) = \beta_0 + \beta_1 \ln(\text{Sales Rank}_j) + \varepsilon_j, \quad (6)$$

Given this specification, β_0 can be interpreted as a measure of the overall demand in the channel, while β_1 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_1 will be less negative in the Internet channel than in the catalog channel (items with lower sales ranks will retain a higher share of channel demand). Brynjolfsson, Hu and Smith (2003) are among the first to use this approach to study the distribution of product sales. They find that such a log-linear curve, also known as the Pareto curve, fits the relationship between product sales and sales rank very well. Chevalier and Goolsbee (2003) also fit sales and sales rank data to a slightly different log-linear relationship with good success. This curve has been successfully used by other economists, who have used it to describe the distribution of income, wealth, and city size (for early examples see: Pareto 1896; and Zipf 1949).

We estimated Equation (6) separately for the Internet and catalog demand data and report both sets of findings in Table 2. When estimating the model we again focused on a random sample of 100,000 transactions 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 results in Table 2 (and those 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_1 coefficients between the two models. The difference in the β_1 coefficients is significant ($p < 0.01$).⁵ This indicates that product sales in the Internet channel are significantly more evenly distributed than product

⁵ The t-statistic of this difference is: $\frac{(\beta_{1I} - \beta_{1C})}{\sqrt{Var(\beta_{1I} - \beta_{1C})}} = 7.779$.

sales for the catalog channel. To the extent that this result reflects lower search costs on the Internet, this finding is consistent with Proposition 2. However, as we recognized in our earlier discussion, the difference in the sales distribution could also be due to the difference in the number of featured products in each channel.

Table 2: 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, $p < 0.05$.

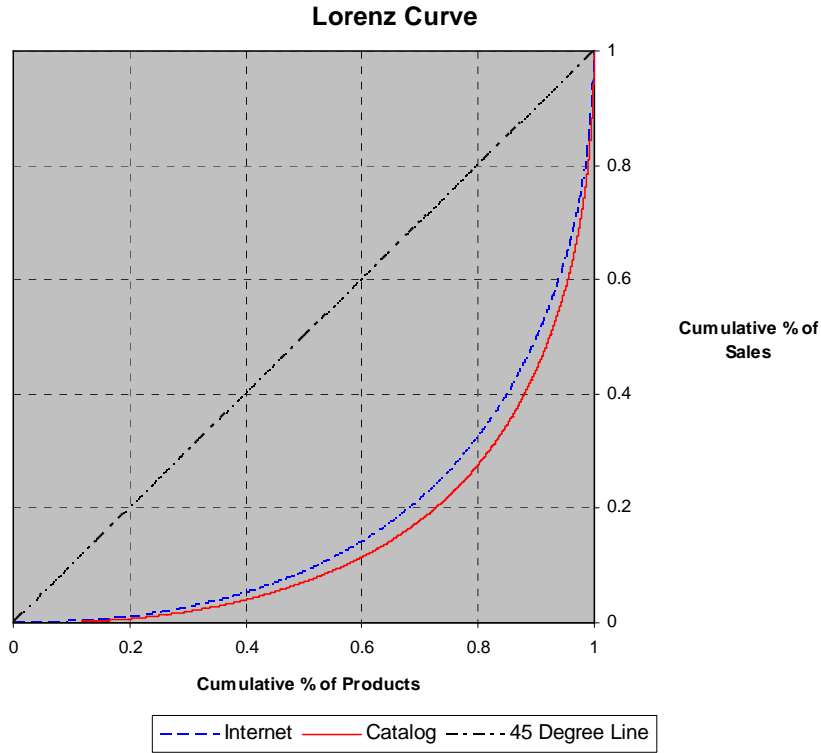
Controlling for Consumer Differences

There are a variety of factors other than lower search costs (and lower featuring costs) that can lead to a more equal sales distribution on the Internet. One obvious explanation is that consumers who shop through the Internet channel are systematically different from consumers who shop through the catalog channel. As a result of this potential selection bias, consumers on the Internet may buy more obscure products than consumers using the catalog channel. Our data provides an opportunity to evaluate this alternative explanation.

In 2002, 220,028 consumers purchased at least one item through the Internet channel, 762,499 consumers purchased at least one item through the catalog channel. These two samples of consumers have an intersection of 53,752 consumers who purchased at least one item through both channels. By focusing on this sub-sample of consumers, we can study how the same set of consumers behaved in the Internet channel as compared to the catalog channel.

Figure 2 presents two Lorenz Curves, when we only consider this sub-sample of consumers who purchased at least one item through both channels. To control for differences in demand, we again focus on a random sample of 100,000 transactions from each channel. Overall, these 100,000 transactions included 5,626 different products in the Internet channel, and 5,474 products in the Catalog channel. The union of these two samples 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 more equal 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 used the 53,752 consumers' transaction data to re-estimate Equation (6) and evaluate how the sales quantity varies with the sales rank across the two channels. We again 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 are reported in Table 3. They replicate the earlier results: the β_1 coefficient is significantly ($p < 0.01$) less negative for the Internet than for the catalog channel, indicating the sales distribution has a longer tail in the Internet channel even when holding the sample of consumers fixed.⁶

Table 3: 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, $p < 0.05$.

After controlling for consumer selection, the difference between the Internet and catalog channels is smaller, as measured by the difference in the β_1 coefficients. The difference between these two channels' Gini Coefficients also decreases when controlling for consumer selection. It appears that consumers who purchase through the Internet channel are different from consumers who purchase through the catalog channel, and this difference is a partial, although not complete, explanation for the difference in the sales distributions.

⁶ The t-statistic of the difference in β_1 is: $\frac{(\beta_{1I} - \beta_{1C})}{\sqrt{Var(\beta_{1I} - \beta_{1C})}} = 4.400$.

Controlling for Unobserved Differences Between the Channels

As we have discussed, the retailer's reliance on a single fulfillment process for both channels provides an effective control for most supply-sided differences, while the analysis of consumers who purchase from both channels controls for consumer differences. However, it is possible that there are other unobserved differences between these two channels, including for example, the cost of featuring products (and the resulting number of featured products). Next we use an individual-level analysis to address this issue.

Instead of studying how sales distributions vary across channels, we study how the distribution of sales varies across consumers with different search costs. In particular, we might expect that search costs on the Internet vary across consumers according to their relative experience with the company's Internet channel. A consumer who has extensive prior experience purchasing over the firm's Internet channel is more likely to be familiar with the site's searching, browsing, and recommendation features. This familiarity may effectively lower their search costs, allowing them to search more quickly through alternative product offers. Thus, we will use a consumer's prior experience with the Internet channel to measure consumers' individual search costs on the Internet and investigate how this correlates with the products that consumers purchase in each channel. If lower Internet search costs lead to a less concentrated sales distribution on the Internet, then we would expect to find that the Internet purchases made by consumers with prior Internet experience are skewed towards more obscure products. As we discussed, because this individual-level analysis focuses on purchases through a single channel, the findings also help to isolate the impact of search costs from the impact of cross-channel differences in the number of featured 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 consumers who do not have prior Internet

experience, we set this variable to zero. Notice that neither measure is defined for new customers who were acquired after January 1, 2002. Therefore we omit new customers from the sample and focus on the 37,835 customers acquired prior to 2002.

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 this consumer's Internet purchases are skewed towards obscure products: the larger the *Average Sales Rank_i* the more obscure the products purchased by customer *i*.

We then regressed 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 compares the percentage change in *Average Sales Rank* for consumers with and without prior Internet experience. Because, a consumer's prior Internet experience may be confounded by the general propensity of the consumer to purchase goods, we also include explicit controls for 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, 2002.⁷ These so-called "RFM" measures are widely used in the catalog industry to segment consumers and provide natural candidates for control variables. This yields the following specification:

$$\ln(\text{Average Sales Rank}_i) = \alpha + \beta_1 \text{Prior Internet Experience}_i + \beta_2 \ln(\text{Recency}_i) + \beta_3 \ln(\text{Frequency}_i) + \beta_4 \ln(\text{Monetary Value}_i) + \varepsilon_i \quad (7)$$

We estimated two versions of the model, replacing *Prior Internet Experience_i* with *Days Since First Internet Purchase_i* (in its natural log) in the second version. The findings are reported in Table 4.

The *Prior Internet Experience_i* coefficient is positive and highly significant. It indicates that the average sales rank of products purchased by consumers with prior Internet experience is about 8.1% larger than

⁷ *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.

that for consumers without prior Internet experience. Similarly, in the second model, the *Days Since First Internet Purchase_i* coefficient is also positive and significant. We interpret these results as evidence that Internet purchases made by consumers with prior Internet experience are skewed toward obscure products. This empirical evidence is consistent with the prediction that consumers with lower search costs have a less concentrated sales distribution.

An alternative explanation for the findings is that customers with more experience have already purchased the popular items, and so these customers only have the more obscure items left to purchase. However, if this was the case we would expect to observe a positive coefficient for the *Frequency_i* variable in the two models in Table. Both *Frequency_i* coefficients are negative and highly significant, indicating that customers who had made more prior transactions were actually *less* likely to purchase obscure items.

Table 4: The Effect of Consumers' Prior Internet Experience on Consumers' Tendency to Purchase Obscure Products through the Internet Channel

	Model 1	Model 2
Intercept	4.913** (0.071)	4.938** (0.070)
Recency	-0.011* (0.006)	-0.014** (0.005)
Frequency	-0.200** (0.013)	-0.202** (0.013)
Monetary Value	0.320** (0.013)	0.319** (0.013)
Prior Internet Experience	0.081** (0.017)	
Days Since First Internet Purchase		0.014** (0.003)
R ²	0.025	0.025
Sample Size	37,835	37,835

Standard errors are in parentheses.

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

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

An even stronger approach for controlling for the unobserved customer differences is to compare whether the relationships reported in Table 4 also extend to the catalog channel. In Table 5 we report the findings

when replicating the analysis using the same sample of consumers (who had purchased from both channels) but focusing on their purchases from the catalog channel. For each of the 37,835 consumers, we identified the products they had purchased through the catalog channel in 2002 and calculated the average sales rank of these products, where the sales rank for each product was calculated using aggregate sales in the catalog channel. We then used this *Average Sales Rank* as the dependent measure in Equation (7). By using sales ranks specific to each channel we control for differences in which products were featured in each channel. However, we obtain a similar pattern of results if we use a common (aggregate) sales ranking across the two channels.

Table 5: The Effect of Consumers' Prior Internet Experience on Consumer's Tendency to Purchase Obscure Products through the Catalog Channel

	Model 1	Model 2
Intercept	5.736** (0.050)	5.717** (0.049)
Recency	-0.022** (0.004)	-0.020** (0.004)
Frequency	-0.205** (0.009)	-0.205** (0.009)
Monetary Value	0.295** (0.009)	0.295** (0.009)
Prior Internet Experience	-0.045** (0.012)	
Days Since First Internet Purchase		-0.007** (0.002)
R ²	0.035	0.035
Sample Size	37,835	37,835

Standard errors are in parentheses.

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

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

The findings are reported in Table 5. The *Prior Internet Experience_i* coefficient remains highly significant, but the coefficient is now reversed, indicating that consumers with previous Internet experience tended to purchase less obscure products through the catalog. Similarly, the coefficient for the *Days Since First Internet Purchase_i* measure is also now negative. We conclude that the contrasting

findings in Tables 4 and 5 suggest that consumers who use both channels tend to use the catalog channel to purchase the most popular products and use the Internet to purchase more obscure items. The contrast in findings between the two channels is precisely the pattern we would expect if search costs are lower on the Internet. Notably, these findings are not consistent with the alternative hypothesis that the change in sales distributions across channels merely results from a selection effect across consumer types.

4. 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 distribution of product sales. This paper provides a theoretical model that investigates how a reduction in search costs affects this distribution. This model also illustrates how a reduction in search costs may affect a multi-product firm's incentives to invest in the marketing of featured products and/or the incentive to introduce new products and varieties.

We present empirical evidence that the Internet channel exhibits a significantly less concentrated sales distribution when compared with the traditional channel. 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 our analysis on consumers who shop through both channels, we are able to control for this potential selection bias. We find that the difference in sales distributions remains statistically significant.

Finally we use an individual-level study to investigate whether and how measures of an individual consumer's search costs on the Internet correlate with which products consumers purchase in each channel. We find evidence that Internet purchases made by consumers with prior experience are more skewed toward obscure 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 customers. It provides

further evidence that lower search costs can lead to a less concentrated sales distribution. Because the underlying technological drivers that we study are certain to continue to progress in advanced economics, the implications for sales distributions, marketing, development strategy 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. 2006. Does Demand Fall When Customers Perceive That Prices Are Unfair: The Case of Premium Pricing for Large Sizes,” *Marketing Science*, forthcoming.
- 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.
- 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. Working Paper. University of Pennsylvania.
- 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>.
- 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.
- Professional Publishing Report. 1999. University Presses Credit Internet for Increased Sales. Volume 3, Number 2, January 29.
- 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 featured 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 making a purchase. 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 making a purchase. Consumer i whose t_i satisfies

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

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

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

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

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

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

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

$(1-\alpha) * (t_i - p_N)$ for $t_i \in [p_N, p_F]$, and 0 for $t_i \in [0, p_N)$. If $s \geq (1-\alpha)(p_F - p_N)$, 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 $t_i > p_F$ will purchase the featured product that gives her the highest value and exit the market. Consumer i whose t_i satisfies $t_i \leq p_F$ will exit the market without making a purchase. If

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

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

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 featured product that gives her the highest value and exit the

market. Consumer i whose t_i satisfies $p_N + \frac{s}{1-\alpha} < t_i \leq p_F$ 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 making a purchase. 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 making a purchase. *Q.E.D.*

Proof of Claim 2.

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

Case A) $p_N < p_F$ 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 featured products is

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

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

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

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

and $p_F^* = 1 - \frac{s}{1-\alpha} + \varepsilon$, $p_N^* \in [1 - \frac{s}{1-\alpha}, 1 - \frac{s}{1-\alpha} + \varepsilon]$, $\pi^* = \frac{L}{4}(1 - \frac{s}{1-\alpha} + \varepsilon - c)(\frac{s}{1-\alpha} - \varepsilon)(1-\alpha^k)$, if

$s \leq \frac{(1-\alpha)(1-c)}{2}$.

Case B) $p_N < p_F$ and $(1-\alpha)(p_F - p_N) \leq s < (1-\alpha)(1-p_N)$. 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 featured products is $D_F(p_F, p_N, s, n, k) = \frac{L}{k}(1-p_F)(1-\alpha^k)$, the demand for each of the

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

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

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

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

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

not exist p_N that satisfies $(1 - \alpha)(p_F - p_N) \leq s < (1 - \alpha)(1 - p_N)$.

Case C) $p_N < p_F$ and $s < (1 - \alpha)(p_F - p_N)$. 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 featured products is $D_F(p_F, p_N, s, n, k) = \frac{L}{k}(1 - p_F)(1 - \alpha^k)\alpha^{n-k}$, the demand for each of the

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

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

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

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

$\pi^* = \frac{L}{4}[(1 - c)^2 - (\frac{s}{1-\alpha} \frac{1 - \alpha^{n-k}}{1 - \alpha^n})^2](1 - \alpha^k)\alpha^{n-k}$, 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_F and p_N that satisfy $s < (1 - \alpha)(p_F - p_N)$.

Case D) $p_N \geq p_F$ 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 featured products is

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

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

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

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

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

Case E) $p_N \geq p_F$ 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 featured products

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

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

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

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

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

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

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

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 $p_F^* = \frac{1+c}{2}$, $p_N^* \in [0,1]$. 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 featured products

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

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 $p_F^* = \frac{1+c}{2}$, $p_N^* = \frac{1}{2} \left(1 - \frac{s}{1-\alpha} + c \right)$. The firm's maximum

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

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

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

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

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

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

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

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

Proof of Proposition 2.

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

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

Thus, the Gini Coefficient is

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

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

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

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

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

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

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

The Gini Coefficient is

$$\begin{aligned}
G(s, n, k) &= 1 - \frac{2 \sum_{j=1}^{n-k} [(n+1-j)D_N^*(s, n, k)] + 2 \sum_{j=n-k+1}^n [(n+1-j)D_F^*(s, n, k)]}{(n+1)[(n-k)D_N^*(s, n, k) + kD_F^*(s, n, k)]} \\
&= 1 - \frac{k+1}{n+1} - \frac{n(1 - \frac{s}{1-\alpha} - c)(\alpha^k - \alpha^n)}{(n+1)[(1 - \frac{s}{1-\alpha} - c)(\alpha^k - \alpha^n) + (1-c)(1-\alpha^k)]}
\end{aligned}$$

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

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

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

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

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

Proof of Proposition 3.

Let $\hat{\pi}^*(s, n, k) = \pi^*(s, n, k) - R(n) - M(k)$. The firm's long-term profit maximization problem is

$$\pi^*(s) = \max_{n, k} \hat{\pi}^*(s, n, k). \text{ If } s > (1-\alpha)(1-c), \hat{\pi}^*(s, n, k) = \frac{L(1-c)^2}{4}(1-\alpha^k) - R(n) - M(k). \text{ In order}$$

to obtain comparative statistics for $n^*(s)$, we take the total differentiation of the first-order-condition:

$$\frac{\partial n^*}{\partial s} = - \frac{\partial^2 \hat{\pi}^* / \partial n \partial s}{\partial^2 \hat{\pi}^* / \partial n^2} = - \frac{0}{-R''(n)} = 0. \text{ If } s \leq (1-\alpha)(1-c),$$

$$\hat{\pi}^*(s, n, k) = \frac{L(1-c)^2}{4}(1-\alpha^k) + \frac{L}{4}(1 - \frac{s}{1-\alpha} - c)^2 \alpha^k (1-\alpha^{n-k}) - R(n) - M(k). \text{ Taking the total}$$

differentiation of the first-order-condition give us:

$$\frac{\partial n^*}{\partial s} = -\frac{\partial^2 \hat{\pi}^* / \partial n \partial s}{\partial^2 \hat{\pi}^* / \partial n^2} = -\frac{2L(1 - \frac{s}{1-\alpha} - c) \frac{1}{1-\alpha} \alpha^n \ln \alpha}{-L(1 - \frac{s}{1-\alpha} - c)^2 \alpha^n (\ln \alpha)^2 - 4R''(n)} \leq 0. \text{ Therefore, } \frac{\partial n^*(s)}{\partial s} \leq 0 \text{ for any}$$

$s \geq 0$. *Q.E.D.*

Proof of Proposition 4.

Let $\hat{\pi}^*(s, n, k) = \pi^*(s, n, k) - R(n) - M(k)$. The firm's long-term profit maximization problem is

$$\pi^*(s) = \max_{n,k} \hat{\pi}^*(s, n, k). \text{ If } s > (1-\alpha)(1-c), \hat{\pi}^*(s, n, k) = \frac{L(1-c)^2}{4}(1-\alpha^k) - R(n) - M(k). \text{ In order}$$

to obtain comparative statistics for $k^*(s)$, we take the total differentiation of the first-order-condition:

$$\frac{\partial k^*}{\partial s} = -\frac{\partial^2 \hat{\pi}^* / \partial k \partial s}{\partial^2 \hat{\pi}^* / \partial k^2} = -\frac{0}{-\frac{L}{4}(1-c)^2 \alpha^k (\ln \alpha)^2 - M''(n)} = 0. \text{ If } s \leq (1-\alpha)(1-c),$$

$$\hat{\pi}^*(s, n, k) = \frac{L(1-c)^2}{4}(1-\alpha^k) + \frac{L}{4}(1 - \frac{s}{1-\alpha} - c)^2 \alpha^k (1-\alpha^{n-k}) - R(n) - M(k). \text{ Taking the total}$$

differentiation of the first-order-condition give us

$$\frac{\partial k^*}{\partial s} = -\frac{\partial^2 \hat{\pi}^* / \partial k \partial s}{\partial^2 \hat{\pi}^* / \partial k^2} = -\frac{-2L(1 - \frac{s}{1-\alpha} - c) \frac{1}{1-\alpha} \alpha^k \ln \alpha}{-L[(1-c)^2 - (1 - \frac{s}{1-\alpha} - c)^2] \alpha^k (\ln \alpha)^2 - 4M''(n)} \geq 0. \text{ Therefore,}$$

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