경영학석사학위논문

The Difference of Optimal Search Set and Effect of Recommendation on Online Shopping – comparing a search goods and an experience goods

온라인 상품 구매 시의 최적 검색량과 추천 효과의 차이 – 탐색재와 경험재의 비교를 중심으로

2016년 2월

서울대학교 대학원

경영대학 경영학전공

김지은

Abstract

JIEUN KIM
College of Business Administration
Seoul National University

Consumers mostly are not aware of all products due to limited recognition capacity. In addition, they search for a limited number of products because of search cost. Search cost can be divided into external cost and internal cost, which includes not only the cost directly related to the search action, but also the cost related to information processing and evaluating products.

This research added additional assumptions to the consumer search model of Kim et al. (2010). Simultaneously considering the limited awareness set and limited search, I estimated each search cost of both coffee makers and whole coffee beans. Consistent with the assertion of Huang et al. (2009), that consumers spend more time to evaluate experience goods, whole coffee beans, which are close to experience goods, have a remarkably higher search cost than that of coffee makers, which are close to search goods. Conducting a simulation study and counterfactual experiment with estimated parameters, experience goods have a smaller optimal search set size. In addition, consumers are more likely to change their choice or buy products when products' recommendation or references are given in the case of experience goods. Therefore, I concluded that accurate product recommendations that suggest a high-utility product to consumers are more important for experience goods.

Keywords: Search Goods, Experience Goods, Search Cost, Limited Search, Structural Modeling, Effect of Recommendation, Random Coefficient Discrete Choice Model

Student Number: 2014-20421

Contents

Tables

Figures

1. Introduction

Online shopping is an important method of purchasing products these days. It decreases search costs, increases consumers' ability to access to products, and increases consumer convenience by lowering the cost of gathering and sharing information. (Hoffman and Novak, 1996, Huang et al, 2009) Its inherent characteristic that consumers cannot directly experience the products before purchase has a significant impact on consumers' search behavior and attitude. This intangibility issue means that consumers search for detailed product information to minimize their uncertainty.

If there is no cost in searching, searching all possible products is the most reasonable and effective strategy. However, search costs play a significant role in real life consumer search and purchase behaviors. (Seiler, 2013)

Search cost includes internal and external cost, and it differs depending on the product category and characteristics. For example, when a product is purchased infrequently, the effort to investigate all products might outweigh the benefit of finding the product that maximizes utility, (Seiler, 2013) which means low search cost. In this research, I have focused on the difference of search cost between search goods and experience goods, and the consequences that follow. Search goods have objective and detailed specifications and their utility is almost entirely explained via specific characteristics, so consumers can get relatively accurate information and feel less uncertainty. In contrast, it is usually difficult to get certain information about experience goods, since their characteristics are difficult to describe and deliver. It is usually combined with specific images or emotional words to help consumers understand the product. In this research, I used experience goods that are relatively easy to divide by their characteristics, which is not necessarily the case for all experience goods. Still, consumers have to contemplate them to evaluate them accurately.

I made a hypothesis that the search cost for experience goods is higher than that of search goods. This means that consumers may search in-depth when they decide to buy experience goods online, to get more accurate information and relieve this uncertainty. In this research, I tried to find the answer the following hypothesis,

1. Experience goods have a higher search cost compared to search goods, so the number of products in its optimal search set is significantly higher than in the search goods optimal search set.

2. The Amazon recommendation system can reduce the search cost and increase consumer utility for both experience goods and search goods.

However, the extent of the increment is significantly different – consumers are more influenced by recommendation and get more utility from it when they search for experience goods, to reduce uncertainty.

In terms of methodology, I employed the model from Kim et al. (2010) to estimate the search cost in an online market. This has two significant advantages compared to other models. First, the model needs only view-rank data, which are open to the public; it does not need either market sales or consumer demographic data. Second, although the model is based on a dynamic optimization problem, it has an analytic solution from the definition of "reservation cost" and "reservation utility," so it does not cause a computation burden or "curse of dimensionality" problem, which are often challenges in dynamic programming. The main difference of this research is that I started from a different assumption. Kim et al. (2010) assumed consumers' full information, which means that consumers know and recognize all products, and limited search, I assumed consumers' limited information and limited search. Second, my research question was mainly based on the difference in the search cost between experience goods and search goods, while Kim et al. (2010) mainly investigated the market structure and the degree of utility increment from choice recommendations.

2. Literature Review

2.1 Search Good vs. Experience Good

Nelson (1970, 1974) classified goods by whether the quality variation was ascertained predominantly by search or by experience, and the respective goods were called “search goods” and “experience goods.” Search goods are defined as those dominated by product attributes for which full information can be acquired prior to purchase; experience goods are dominated by attributes that cannot be known until the purchase and use of the product or for which an information search is more costly and/or difficult than direct product experience. (Klein, 1998) Their difference is significant in advertising. While advertising for search goods provides direct information to the consumers about the qualities of a particular goods, there is little direct information contained in the advertising for experience goods. (Leahy, 2005) In the situation of searching, Huang et al (2009) showed that consumers spend more time per product in the case of experience goods on online, because they require more time to assess quality.

2.2 Limited Awareness Set

Consumer choice is usually influenced by their awareness set. Starting Howard and Sheth (1969) who suggested the concept of 'evoked set', their recognition and selection processes have been studied before. A two-stage process model explains their consideration process when making a purchase; the consumer might undertake a two-stage process, first filtering the available alternatives and then undertaking a detailed analysis of the reduced set (Wright and Barbour 1977). Limited consideration set is related to not only cognition capacity but also evaluation cost. Large consideration set needs more cost to evaluate each alternative.

Bronnenberg (1996) found that ignoring a limited choice set may result in biases in price response and price competition. However, a traditional choice model, such as in Berry, Levinson, and Pakes (1995), mainly assumes that consumers have full information about the products that they are considering buying. It means that consumers know the characteristics of each product, so they are able to find what can maximize their utility by comparing the utility of each product. Nevo (2001) extended this choice model and investigated market structure and firms' pricing behavior. Goolsbee and Petrin (2004) estimated the nature of competition between Direct Broadcast Satellites (DBS) and cable TV by investigating price elasticity and price response using a probit model. All of them assumed homogeneity in the consumers' consideration set. Consideration set heterogeneity, which is due to limited information or information asymmetry between consumers, was included by Goeree (2008), who assumed that each consumer has different information, according to their degree of exposure to media and product advertising, and termed this information heterogeneity. She indicated that ignoring information asymmetries between consumers and firms results in biased demand curves. Draganska and Klapper (2011) used microlevel tracking data and found that considering a heterogeneous choice set improves estimation accuracy. They explicitly considered consumer heterogeneity in brand awareness and improved their specification and estimation of aggregate discrete choice models of demand and augmented consumer tracking data with sales and marketing-mix data.

2.3 Consumer Search

Consumer search is studied and adopted to make search cost estimations (Seiler, 2013), sponsored search advertising effects, and bidding strategies (Yao and Mela, 2010, Chan and Park 2015). Articles related to this topic all explicitly considered search costs; the cost always matters because if there were no search costs, consumers could find all information and make fully informed choices. Furthermore, Moraga-Gonzalez et al. (2007) suggested a semi-parametric estimation methodology to estimate the search cost under a non-sequential search assumption. Search cost is important factor in consumer search process. Since consumers have to pay search costs such as time or cognition, their search behavior is limited to within several products. Therefore, searching cost heterogeneity leads to consumer choice set heterogeneity, and they have to make decisions with limited information.

Search cost is set in cost/benefit framework. Here, benefit of a point of time, as it is mentioned in the following section, implies maximum utility of products that have searched until then. Search cost can be divided into external cost and cognitive cost. External search cost is the direct cost of resources buyers invest in search, such as time and money, and also includes opportunity costs of them in foregone search activities. (Smith et al., 1999) Also, cognitive costs, also known as internal cost, means cognitive effort that needs to sort information and integrate them to make decisions. (Hauser et al., 1993, Smith et al., 1999) Hauser and Wernerfelt (1990) established evaluation cost model in the same vein with search theory and cost/benefit frame. Roberts and Lattin (1991) assumed direct experience and included mental maintenance and processing cost from experience in search costs of packaged goods. In contrast, Kim et al. (2010) interpret search cost as the opportunity cost of time invested in identifying and evaluating another candidate product in the context of durable goods. However, direct experience is impossible on online environment, so experience goods is evaluated based on description of products. Therefore, I interpret search cost same in search goods and experience goods, as it include browsing cost and cognitive effort.

Consumer searches can be roughly classified into two cases: Attributes search and price search by object of search, and sequential search and non-sequential search by method of search. In the attribute search process, consumers make uncertain parts of attributes certain through searching. In the price search process (Mehta et al. 2003, Hong and Shum 2006), consumers face price uncertainty and find a price (in most cases, the lowest one) of products that have specific characteristics. In particular, Hong and Shum (2006) developed a methodology to estimate the search cost from price data under the assumption that each price is an equilibrium price resulting from each consumer's optimal search.

In sequential search, a consumer weighs the expected benefits and costs of gathering additional product information after each new piece of product information has been updated. Meanwhile, in non-sequential, simultaneous search, the consumer samples a fixed number of alternatives and purchases the alternative with highest utility in this set. (Baye et al. 2006, Hong and Shum 2006, Wildenbeest 2006, Honka and Chintagunta, 2013) The example of non-sequential search is in the automobile market where consumers make appointments with a number of dealers beforehand. (Moraga-Gonzalez et al, 2015) Honka and Chintagunta (2014) compared sequential and non-sequential search models in the same industry and concluded that the better model depends on the size of the company. Sequential search assumption is usually accepted in the online market. Weitzman (1979) characterized the solution of sequential search. Kim et al (2010, 2015) and Chan and Park (2015) assumed sequential search behavior for modeling consumers online. Kim et al. (2015) added consumers' real purchases to a model from previous research.

Attributes	Ranges
Brand	Black_decker (13), Cuisinart (23), Hamilton (12), Mr.coffee (31), Other (16)
Size	Small (9), Big (61), Middle (25)
Programmable	Yes (63), No (32)
Thermal	Yes(38), No (57)
Color (can be counted twice)	Black (57), White (19), Silver (21), Other (8)
Price	$55.135 (mean), $43.331 (std. dev.)

Table 1 Description of the Choice Options in Coffee Maker

Attributes	Ranges
Brand	Coffee Bean Direct (8), Café Don Pablo (6), Eight O' Clock (8), Fresh Roasted Coffee (28), Kicking Horse (3), Koffee Kult (3), Lavazza (11), Other (29)
Roast	Dark (19), Other (medium ,light ,unknown) (76)
Volume	2 lb. (52), 5 lb. (17), Other (27)
Decaffeinated	Yes (16), No (80)
Organic	Yes (25), No (71)
Sour	Yes (10), No (86)
Price	$22.942 (mean), $10.302 (std. dev.)

Table 2 Description of the Choice Options in Coffee Bean

3. Data

I selected two products in each category for analysis, and I chose “whole coffee bean” in the experience goods category and “drip coffee maker” in the search goods category. The reasons for this are|: First, these two kinds of products are in the same category and have similar usage, resulting in possible bias from different traits, which consumption recognition can reduce. Second, both of them have characteristics that can be described and divided into several factors. Experience goods are usually demonstrated by its overall image or in combination with other subjects, so it is difficult to adopt an econometric model to analyze it. However, whole coffee beans can be categorized with several criteria, such as country of origin and degree of roasting. These categories give consumers overall expected quality information about the coffee beans, but the information is relatively incomplete and leads to an indeterminate comparison with normal search goods. Third, both products belong to one of certain categories for which Amazon.com provides view-rank information. In the “kitchen & dining” category, Amazon provides information to customers for each product, such as view-list (“Consumers who viewed this product also viewed…”), buy-list (“Consumers who bought this product also bought…”) and similar products recommendation based on product specification and consumer search and purchase records. Although the rank is truncated, it still gives plenty of information about consumers' search priorities. Fourth, these products contain product information on their name. Therefore, consumer can know its value before they click to search, which consistent with model assumption.

For selecting individual products, I first referred to the top 150 products from the best-selling ranking. After that, I removed several products that are rarely mentioned from the view-rank list or if their view-rank lists consisted of a large portion of outside goods. I finally selected 95 drip coffee makers and 96 types of whole coffee beans. Since product names contain information about products, I categorized them according to product information presented on their name. Some coffee makers have both black and silver color on the same body. In that case, I classified them into both black and silver. When it comes to coffee bean, I classified dark, Italian, French roasted coffee into 'dark', and coffee bean originated from Costa Rica and Ethiopia into 'sour'.

I collected data from the Amazon website from September to November. To minimize temporal demand shock and idiosyncratic errors, I avoided using data collected at one time point. Instead, I aggregated the view-rank list for two weeks conditional on each product in a daily basis. I averaged a daily rank for each product and ordered them in a descending order. I gave the same aggregated rank with the products of the same average.

I also used the 24 best-selling product list that appeared before filtering, and the first 24 products list after I filtered out by each criterion. These are mainly related to consumer recognition set, which will be explained later. To specify search cost, I chose two variables for each products. In coffee maker, I used the recommendation list and the list of products that were shown on the first page of view-list. In coffee bean, I used the list of products that were shown on the first page of view-list and buy-list. I considered that Amazon.com does not provide recommendation of whole coffee bean and there is almost no coffee maker on buy-list of coffee maker. These references function as shortcuts from one to the other. A direct connection to another product's influence on consumer choice by decreasing their search cost.

4. Model

4.1 Utility and Search Cost

The model I used is mostly same as Kim et al. (2010). I assume individual heterogeneity for each product. Individual consumer utility for individual product can be represented as,

$$u_{ij} = V_{ij} + e_{ij}, \qquad e_{ij} \sim N(0, \sigma_{ij}^2)$$

where e_{ij} is idiosyncratic term of utility which are not known to consumer in advance. Most literature dealing with market share data assumes that e_{ij} follows Type 1 extreme value, since it has a convenient integral form for calculating the consumer choice probability. However, there is no information about consumer choice probability in the Amazon website, so a flexible form is more appropriate here. Furthermore, due to the identification problem, I ignore unobserved product quality data (unobserved by the researcher, observed by consumers). Amazon.com provides a brief explanation about the traits of each product in their name. Therefore, prior to searching, a consumer is able to form an expectation by calculating the value, V_j, of each product through its name. After clicking on the product, they can learn its entire utility because they can find e_j.

4.2 Sequential Search

In a sequential search process, consumers continue to search when they expect that the marginal benefits from additional searching exceed its marginal costs. A search decision is made in every searching process, in contrast to a non-sequential search, for which the search decision is made based on accumulated search results.

Sequential search is an inherently dynamic decision process in which each agent considers future rewards. A consumer forms an expectation about options not searched, and compares the expected utility of search costs to make a decision about whether to search. In other words, they maximize the current utility considering future rewards.

Weizman (1979) expresses each search pause with a bellman equation as follows:

$$\psi(\bar{S}, y) = \max\{ y, \max_{i \in \bar{S}}\{-c_i + \beta_i[\psi(\bar{S} - \{i\}, y) \int_{-\infty}^{y} dF_i(x_i)$$

$$+ \int_{y}^{\infty} \psi(\bar{S} - \{i\}, x_i) dF_i(x_i)]\}\}$$

Here, $\psi(\bar{S}, y)$ denotes the expected present value of the following optimal policy under the state, $\bar{S}$ and y are state variable, $\bar{S}$ the maximum reward from previous choices. If a consumer chooses one product, only the highest utility matters. c_i indicates search cost, which is different for each product. β_i is the discount rate due to the time interval between two consequent searches.

The optimal sequential search process has an optimal stopping point, since the consumer continues to search only if the expected marginal utility is higher than a certain point. The point makes the consumer indifferent between searching and stopping. It is called "reservation utility," which is denoted z_i. Therefore, it is an optimal stopping problem, similar to Rust (1987).

In a sequential searching process, only the highest utility among the utility that has already been found matters, because searching activity requires a searching cost, and it is more reasonable to stop searching if the consumer finds another product that is expected to have a higher utility. Define u_i^* as the highest utility among the products so far searched, and consumer i's expected marginal utility $\mathcal{B}_{ij}$ from additional searching j is

$$\mathcal{B}_{ij}(u_i^*) = \int_{u_i^*}^{\infty} \left(u_{ij} - u_i^*\right) f(u_{ij}) du_{ij}$$

Where $f(u_{ij})$ is the probability density function of u_{ij}. If u_{ij} is lower than u_i^*, the marginal utility does not change. Meanwhile, when u_{ij} is higher than u_i^*, the marginal utility changes to u_{ij}. The consumer continues to search when the expected marginal utility is higher than the search cost.

In Amazon.com, a consumer is faced with a list of products, and investigates each product by clicking on its name. Consumer i's dynamic decision process can be expressed as

$$W(u_i^*, \bar{S}_i) = \max\left(u_i^*, \max_{j\epsilon\bar{S}}(-c_{ij} + \beta_i \cdot [\, F(u_i^*) \cdot W(u_i^*, \bar{S}_i - \{j\})\right.$$

$$+\int_{u_i^*}^{\infty} W(u_i^*, \bar{S}_i - \{j\}) f(u_{ij}) du_{ij}]))$$

where $\bar{S}_i$ is the set of products not searched and j denotes each product.

The condition that the consumer keeps searching is

$$c_i < \mathcal{B}_{ij}(u_i^*)$$

The reservation utility, z_{ij}, is the threshold that consumer decides to stop searching. Since the expected marginal utility of z_{ij} makes the consumer indifferent to continuing searching, it can be written as

$$c_i = \mathcal{B}_{ij}\left(z_{ij}\right) = \int_{z_{ij}}^{\infty} \left(u_{ij} - z_{ij}\right) f(u_{ij}) du_{ij}$$

The optimal sequential search strategy proposed by Weitzman (1979) consists of two rules:

(1) Selection rule: A consumer searches for a product with the highest reservation utility among the products in their available or consideration set.

(2) Stopping rule: A consumer terminates their search process when the maximum utility that they have obtained exceeds the reservation utility of all unsearched products, which means that the highest utility matters.

If the consumer only chooses one product, they collect maximum utility, u_i^*

by choosing the product that gives u_i^*.

Next, in the following equation, $r(j)$ denotes the product with jth highest reservation utility, and $\pi_{i,r(j)}$ is probability that product with jth highest reservation utility is searched. I assumes that $e_{i,r(k)}$ follows the standard normal deviation, then $\pi_{i,r(j)}$ can be calculated as

$$\pi_{i,r(j)} = \Pr\left(\max_{1\le k\le j-1}(V_{i,r(k)} + e_{i,r(k)}) < z_{i,r(j)}\right)$$

$$\Pr\left(\bigcap_{k=1}^{j-1}\left\{e_{i,r(k)} \le z_{i,r(j)} - V_{i,r(k)}\right\}\right)$$

$$= \prod_{k=1}^{j-1} F(z_{i,r(j)} - v_{i,r(k)})\,, \qquad j > 1$$

Based on optimal search process, $\pi_{i,r(j+k)}(k, j > 0)$ is always lower than $\pi_{i,r(j)}$. Moreover, a consumer who searches for products with $(j + k)$th highest reservation utility also search that with highest jth

$$\pi_{i,\{r(j) and\ r(j+k)\}} = \pi_{i,r(j+k)} = \min\left(\pi_{i,r(j)}, \pi_{i,r(j+k)}\right)$$

I assumes that consumers use the filtering that Amazon.com provides and that their consideration sets are limited. In other words, consumers do not consider the reservation utility of all coffee makers. Instead, they are able to recognize some part of the products due to their limited recognition capacity and their tendency to improve their search efficiency. The limited-capacity model of attention was developed to explain these selective and intensive aspects of attention (Kahneman 1973). This model assumes that one’s total attentional capacity at any one point in time is limited.

Kim et al. (2010) assumes that consumers recognize all products and calculate the reservation utility. In most cases, they do not browse and recognize all listed products. Instead, they first take notes of the products on the first page. Furthermore, since Amazon.com enables consumers to filter products by certain criteria, reasonable consumers use it to improve the effectiveness of the search process. They filter products before searching and make their decisions (Yao, Mela 2010). So I added following assumptions,

Assumption 1. The consumer always recognizes the 24 best-selling products at the first page before filtering.

Assumption 2. The consumer filters products and mainly considers the products with the characteristic to which they are most sensitive. The consumer always recognizes the 24 products on the first page after filtering.

Assumption 3. The consumer can put other products in their consideration set through other parts of Amazon.com page, and they are selected randomly.

Thus, each consumer's consideration set in the model contains:

- A consumer filters products according to their highest brand and feature coefficient.

- A consumer always browses the first page, so the first page products before and after filtering are always in the consideration set.

Amazon.com provides the view-rank list of each focal product in order of decreasing prevalence for several product categories. According to the Amazon.com U.S Patent 6,912,505 (Linden et al. 2001), a commonality index, CI_{jk}, defined as

$$CI_{jk} = \frac{n_{jk}}{\sqrt{n_j}\sqrt{n_k}}$$

where j, k indicate each product, n_j means the number of people who clicked product j, and n_{jk} means the number of people who click both j and k. As the formula shows, CI measures the degree of the strength of the relationship between two products. When many consumers click on both of them, their CI is high, which means that two products have a close relationship, such as similarity in characteristics and product image, or they give consumers a similar amount of utility. In contrast, when few consumers click both and click only either j or k, CI is low, which means that they have little similarity. $n_j, n_k \geq n_{jk}$ so commonality index is bounded between 0 and 1. Based on view-rank list provided by Amazon.com, an indicator variable can be defined as

$$I_{j,kl} = \begin{cases} 1 & \text{if } (j,k) > (j,l) \\ 0 & \text{otherwise} \end{cases}$$

I can compare view-rank between listed products directly. Listed products automatically have a meaning that they have higher probability to be searched than unlisted products. Although I cannot decide rank between unlisted products, the above comparison also contains enough information.

5. Estimation

5.1 Specification

I used a random coefficients discrete choice model to calculate a consumer's product search utility. I did not include product characteristic terms that were only observed by the consumer and not observed by the researcher or product-specific fixed effect term, since I only have rank data, meaning that these terms were not directly identified. The utility of product j for consumer i can be expressed as

$$u_{ij} = X_j\beta_i - \alpha_i P_j + e_{ij}$$

Where X_j is a $k \times 1$ row vector of product j's characteristics, P_j is its price, α_i represents the individual-specific price sensitivity, and β_i indicates a $1 \times k$ column vector of individual specific sensitiveness to each product characteristics. e_{ij} is individual and product specific idiosyncratic error term. I assume α_i, β_i, and e_{ij} are normally distributed. To be specific,

$$\alpha_i \sim N(\beta_p, \sigma_p^2), \qquad \beta_i \sim N(\beta_0, \Sigma_\beta), \qquad e_{ij} \sim N(0, \sigma_{ij}^2)$$

where Σ_β is a diagonal matrix.

Amazon.com provides shortcuts to find particular products; first, Amazon.com recommends three "similar" coffee makers for each product, based on their specified characteristics and consumer view log. In addition, products that are in the front position in the view-rank list can lead consumers directly to click them. To keep search costs larger than zero consistent with the theory, I specified search cost of coffee maker as

$$c_{ij} = \exp\left(\gamma_{0i} + \gamma_{1i} sim_j + \gamma_{2i} frv_j\right),$$

$$\gamma_{0i} \sim N\left(\gamma_0, \sigma_{\gamma 0}^2\right), \qquad \gamma_{1i} \sim N\left(\gamma_1, \sigma_{\gamma 1}^2\right), \qquad \gamma_{2i} \sim N\left(\gamma_2, \sigma_{\gamma 2}^2\right)$$

where sim_j means the frequency of similar product recommendation to product j provided by Amazon.com, frv_j is the appearance frequency of view-rank list in the first page to product *j*. Coefficient terms reflect consumer-specific sensitivity to search cost, and random effects indicate different search behavior. For instance, consumers who count on their own search ability and judgment will rely less on recommendation from Amazon or others' search history. In contrast, other consumers may prefer to take account of what other search and buy, or other reliable recommendation to reduce risk due to the uncertainty on online shopping. Similarly, search cost of coffee maker is

$$c_{ij} = \exp\left(\gamma_{0i} + \gamma_{1i} frb_j + \gamma_{2i} frv_j\right),$$

$$\gamma_{0i} \sim N\left(\gamma_0, \sigma_{\gamma 0}^2\right), \qquad \gamma_{1i} \sim N\left(\gamma_1, \sigma_{\gamma 1}^2\right), \qquad \gamma_{2i} \sim N\left(\gamma_2, \sigma_{\gamma 2}^2\right)$$

where frv_j is the appearance frequency of buy-rank list and frv_j is the appearance frequency of view-rank list in the first page to product *j*.

5.2 Approach and Computational Details

For estimation, the commonality index has to be redefined in a form that the researcher can know.

The commonality index between products j and k can be estimated as

$$\widehat{CI}_{jk}(\theta; \mathrm{X}) = \frac{\hat{n}_{jk}}{\sqrt{\hat{n}_j}\sqrt{\hat{n}_k}}$$

where $\hat{n}_j$ is the forecasted number of individual who search for product *j*. $\hat{n}_j$ can be approximated as;

$$\hat{n}_j = \sum_{i=1\ldots J} \pi_{ij}$$

$$\hat{n}_{jk} = \sum_{i=1\ldots J} \min(\pi_{ij}, \pi_{ik}) \qquad j, k = 1, \ldots, J$$

The estimated commonality index can be decomposed as a true commonality index term and error term:

$$CI_{jk} = \widehat{CI}_{jk}(\theta; \mathrm{X}) - \epsilon_{jk}, \qquad \epsilon_{jk} \sim N(0, \frac{v^2}{2})$$

where ϵ_{jk} is aggregate-level prediction errors. Possible error sources include aggregation or sampling error by the researcher or Amazon, and measurement errors due to hedonic browsers that do not act consistent with their utility. A higher rank means higher commonality index,

$$\Pr(I_{j,kl} = 1) = \Pr(CI_{jl} < CI_{jk}) = \Pr(\epsilon_{j,kl} < \widehat{CI}_{jk} - \widehat{CI}_{jl}),$$

where $\epsilon_{j,kl} = \epsilon_{jk} - \epsilon_{jl}$. Given the assumption about the distribution of ϵ_{jk}, $\epsilon_{j,kl}$ is normal random variable with a mean of 0 and a variance of v^2. Therefore,

$$\Pr(I_{j,kl} = 1) = \Phi\left(\frac{\widehat{CI}_{jk}(\theta; \mathrm{X}) - \widehat{CI}_{jl}(\theta; \mathrm{X})}{v}\right)$$

where Φ is the cumulative density function for the standard normal distribution.

The nonlinear least square estimator to find parameters above model is defined as,

$$\{\Theta^*, v^*\} = arg \min_{\{\Theta^*, v^*\}} \sum_{(j,k,l)\in S} \left[\Pr(I_{j,kl} = 1 - 1)\right]^2$$

where $j, k, l = 1, \ldots, J$ and $\mathrm{S} = \{(j, k, l) | I_{j,kl} = 1, j \neq k \neq l\}$.

The reservation utility can be expressed as,

$$z_{ij} = V_{ij} + \zeta\left(\frac{c_{ij}}{\sigma_{ij}}\right) \times \sigma_{ij}$$

where $\zeta(x_{ij})$ solves the following implicit equation:

$$x = (1 - \Phi(\zeta))(\lambda(\zeta) - \zeta)$$

where λ is the standard normal hazard rate, $\frac{\phi(\zeta)}{(1-\Phi(\zeta))}$.

A state transition matrix or transition equation is usually required to solve a dynamic optimization problem, and it often causes the "curse of dimensionality," mainly due to the dimension of state space. In this article, I cannot explicitly define state transition since I do not have consumers' search records by time. Kim (2010) suggests a solution to solve this problem using the definition of reservation cost.

I used the spline method to interpolate the reservation utility, z_i. First, I generated uniformly spaced points and calculated their corresponding value by minimizing the squared-error of the equation. Next, I set the individual search cost of each product as a query point and interpolated the cost-reservation utility function.

For optimization, I used a hybrid genetic algorithm and pattern search algorithm. There are two sources of the discontinuity of object function, first, "minimum" operator generates a discontinuity in the objective function; second, consumers' sorting behavior by reservation utility also generates discontinuity, because the reservation utility of each product is discrete, and a product's change of utility results in a change of the sequences. Most algorithms that depend on starting values end to find the local minimum. To find a good starting value, I first used a genetic algorithm, a kind of evolutionary algorithm. Next, I obtained a good initial parameter using the Nelder-Mead simplex algorithm. Finally, inputting the previous parameter as a starting point, I use an effective pattern search algorithm to find the optimal point of the discontinuous function, to find the final optimal parameter. I repeat this process several times to find the global minimum.

I used bootstrap samples to compute the standard error. There were two reasons for this decision. First, only one independent sample was used here. Because the view-ranks were dependent on each other, each indicator variable could not be regarded as a sample, as the samples must be independent. Having no information about the population distribution, and being unable to rely on the law of large numbers directly, I had to create simulated samples from the "simulated population." The second reason I chose to use bootstrap samples to compute the standard error had to do with the complex discontinuity of the objective function, which made it almost impossible for me to compute the standard error using widely known gradient methods. The bootstrap technique was first introduced by Efron (1979), then expanded upon by Efron and Tibshirani (1993). It is commonly used to compute the standard error because its computation is very simple and because it can be used regardless of the population distribution. To compute the standard error, I drew a random sample of the same size from a pairwise view-rank indicator with a replacement. Next, I used this sample to estimate the model parameters. The standard error of the parameters was

$$\widehat{SE}_B = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}\left\{\hat{\theta}^*(b) - \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}^*(b)\right\}^2}$$

where $\hat{\theta}^*(b)$ is the estimated parameter with a bootstrap sample and B is the total number of bootstrap samples.

6. Result

6.1 Parameter Estimates

Search goods (coffee maker)	Mean effect (std. error)	Heterogeneity (std. error)
Black_decker	0.847(0.772)	2.332 (0.291)
Cuisinart	-2.087 (0.656)	2.332 (0.291)
Hamilton	2.684 (0.817)	2.332 (0.291)
Mr. coffee	1.546 (0.960)	2.332 (0.291)
Small	0.963 (0.450)	4.726(0.226)
Big	2.387 (1.279)	4.726(0.226)
Programmable	-0.234 (0.581)	0.582 (0.658)
Thermal	1.633 (0.667)	3.096 (0.623)

Black	2.428 (1.004)	2.925 (0.431)
White	0.581 (0.257)	2.925 (0.431)
Silver	-0.272 (0.841)	2.925 (0.431)
Log(Price)	1.971 (0.676)	6.978 (0.088)
Search base cost	-7.550 (1.381)	1.078 (0.545)
Effect of recommendation	-1.422 (0.362)	0.856 (0.388)
Effect of top view-rank	-0.866 (0.091)	0.904 (0.368)
Standard deviation of CI	0.106 (0.005)	-
Number of inequalities	217394	
Sum of squared errors	25396.39	

Table 3 Estimation Results – Search Goods

Experience goods (coffee bean)	Mean Effect (std.error)	Heterogeneity (std.error)
Coffee Bean Direct	1.114(1.498)	4.724(1.106)
Café Don Pablo	3.850(1.565)	4.724(1.106)
Eight O' Clock	-3.124(1.126)	4.724(1.106)
Fresh Roasted Coffee	2.066(0.965)	4.724(1.106)
Kicking Horse	3.789(0.243)	4.724(1.106)
Koffee Kult	2.311(0.382)	4.724(1.106)
Lavazza	-3.446(1.392)	4.724(1.106)
Dark Roast	3.391(1,676)	1.807(0.721)
2-pound	-3.670(0.007)	2.095(0.766)
5-pound	-4.059(1.594)	2.095(0.766)
Decaffeinated	1.630(0.675)	4.730(1.038)
Organic	1.697 (0.950)	3.005(1.010)
Sour	2.025 (0.673)	4.027(1.509)
Price	2.996(1.019)	4.801(1.863)
Search base cost	-2.426(0.987)	0.499(0.330)
Effect of top buy-rank	-0.550(0.195)	0.062(0.035)

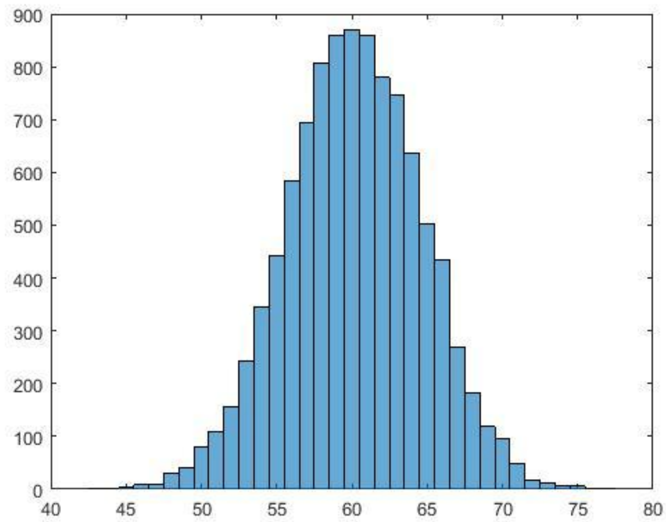
Effect of top view-rank	-0.407(0.190)	0.031(0.024)
Standard deviation of CI	0.146(0.018)	-
Number of inequalities	165013	
Sum of squared errors	25290.90	

Table 4 Estimation Results – Experience Goods

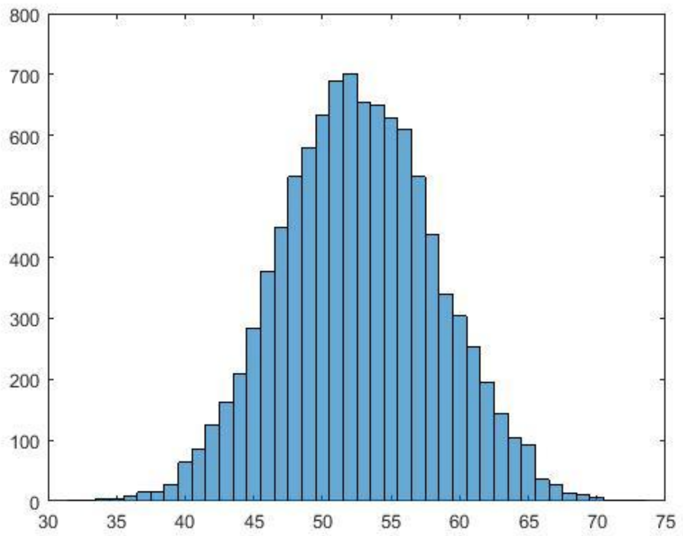
Table 3 and Table 4 show the estimation results for each parameter. The level of consumer heterogeneity toward the brand was found to be high in both products. High heterogeneity indicates that a brand can be highly preferred by some consumers and rejected by other consumers at the same time. The price elasticity of both products was negative, but it is positive for some consumers; this means that some consumers prefer more expensive products because they regard these products as having a higher level of quality than lower-priced items. The results indicated that there are significant differences in base cost of the product and the effect of the references between search goods and experience goods. These factors both indicate congruency between the list of products and the products on the purchase page. It is difficult for consumers to evaluate experience goods based solely on the product description found on a webpage, especially in comparison to search goods. When consumers research different kinds of experience goods, they must imagine themselves experiencing the product; for example, they may imagine themselves smelling or drinking a specific type of coffee. This is a time-consuming process that requires more energy than is needed to research search goods, which creates a difference between the two types of goods in the marginal search cost of one unit of addition introduction of product. The purchase of experience goods also relies heavily on individual preferences and previous experience with the product. Product quality and technological level of the product are also important factors that consumers consider when purchasing search goods, but these characteristics are especially important for experience goods, as it may not be proper to discuss the superiority of quality toward these products. If consumers already know what they like, they will have little motivation to do more search to find better prices. This results in a difference in search-base cost between search goods and experience goods, meaning that consumers pay the same amount regardless of the amount of searches they conduct.

6.2 Robustness

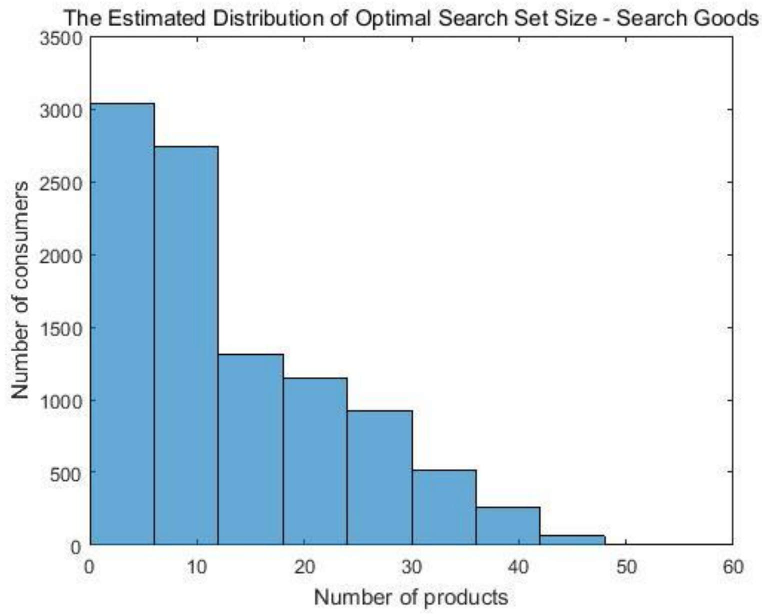
Before estimation, I assumed that consumers would be just as likely to recognize or not recognize a product that was not found on the first page of the search results. I tested the robustness of this theory by assuming different probabilities for recognition and consideration of a product. I changed the probability to 30% (low level) and 70% (high level), which means there are different levels of consumer knowledge about product existence and the number of products in the initial consideration set. After changing the probability, I re-estimated the model parameters and compared them with the original parameters. All parameter estimates showed high correlation. The coffeemaker showed correlations of 0.9054 (30%) and 0.9379 (70%) each, and the coffee beans showed correlations of 0.8164 (30%) and 0.9159 (70%) each. Based on these results, I concluded that the parameter estimates were robust enough to act as alternative assumptions for setting different probabilities.



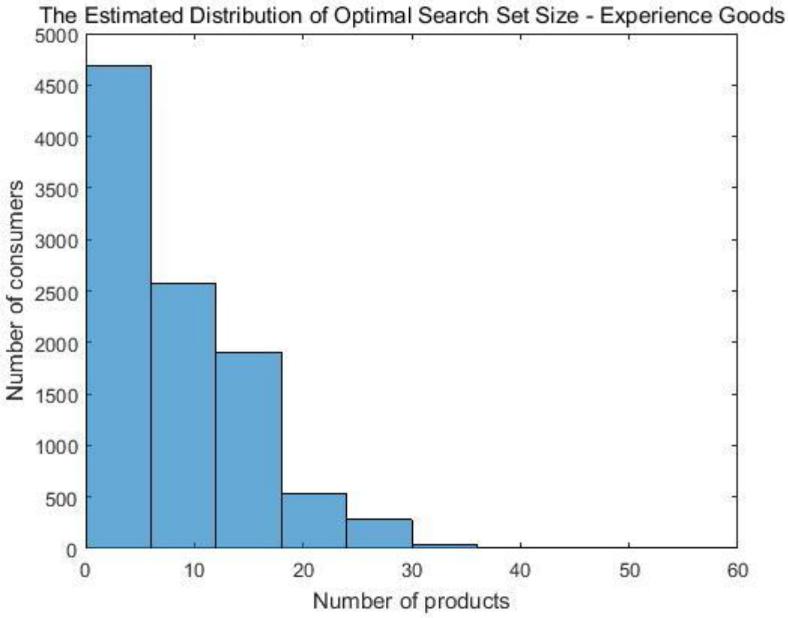
[Figure 1] Total Number of Simulated Consideration Set – Search Goods



[Figure 2] Total Number of Simulated Consideration Set – Experience Goods



[Figure 3] Number of Products in Simulated Optimal Search Set – Search Goods



[Figure 4] Number of Products in Simulated Optimal Search Set – Experience Goods

6.3 Search Set Analysis

In order to verify the effect of search cost on consumer search behavior, I analyze the optimal search set size of each product. I generate 10,000 pseudo-households from the population with estimated parameters. Next, I calculate V_{ij} and Z_{ij} and counted how many V_{ij} are higher than Z_{ij}, based on the model assumption. Since consumers can know the most part of product utility, V_{ij}, before click product to search, I expect that their search set would contain just several products.

Figure 1 and figure 2 show the distribution of the optimal search size more specifically for the convenience of comparing the two products. Due to the difference in search cost, consumers' search set size in search goods is larger on average. In the coffee maker category, the average number of products in the search set is 12.78, the median number of products is 9, and the standard deviation is 10.24. In the coffee bean category, the average number of products is 7.81, the median number of products is 6, and the standard deviation is 6.64. From t-test, I found that these samples are chosen from separated population groups. ($p<0.0001$)

Compared to the consideration set size, which had an average of 55 to 60, the results clearly showed that the search cost affects consumer search behavior. This means that consumer search behavior represents a limited search rather than a full search. Additionally, consumers search more when they are planning to purchase search goods, due to the lower search costs associated with these goods. This also occurs because search goods tend to have more reliable and informative descriptions, making consumers more motivated to search more for these products. The specifications of search goods are relatively easy to explain using written descriptions. These descriptions allow consumers to calculate the expected utility of each product and compare it to that of other products. Consumers also tend to be more mindful when purchasing search goods due to the high prices of these products and the low frequency at which they purchase such goods. With experience goods, on the other hand, it is difficult for consumers to get a full and objective idea of the product through descriptions alone, since the quality of these goods depends strongly on consumers' direct use or on the sensory experience had by users of such products. Therefore, when consumers buy experience goods online, they tend to buy things they have bought before rather than searching for new products.

7. Counterfactual Experiments

As mentioned earlier, Amazon provides shortcuts for searching. For each product, it provides a buy-list of other consumers who have searched for the product. For some products, Amazon provides recommendations based on consumer logs and product characteristics. It sometimes shows the top three products that consumers have actually bought after they searched for certain products or view-lists that consumers have searched for after they searched for the products. In the kitchen and dining category, Amazon provides both the view-lists and buy-lists of others. In the case of coffee makers, Amazon provides recommendations, while it does not in the case of coffee beans.

However, a buy-list does not function as a reference, since consumers seldom buy two coffee makers at the same time, so there are no coffee makers within the first several products of the buy-list. In contrast, consumers often buy two or more types of beans at the same time, so the buy-list functions as a reference as well in the coffee bean page. To function as "shortcuts," references should be located in a conspicuous place. Here, I assumed the top six products on the buy-list and view-list, which are shown to consumers first on the purchase page, could be regarded as substantial shortcuts.

Shortcuts can reduce search cost by recommending or suggesting similar products that are expected to be searched for or preferred by consumers. However, if they lead consumers to the wrong page, even though they reduce search cost, they decrease consumer utility as well, which leads to a decrease of consumer surplus (i.e., consumer utility minus search cost). I conducted several counterfactual experiments to confirm the effect of several shortcuts on web pages and compare the effects of each shortcut path. Although shortcuts are an efficient way to reduce search cost, the extent of the reduction can differ by category. In addition, recommendations from Amazon's own analysis and consumer search lists may have different effects due to differences in source and reliability.

The net surplus of consumer i who has search set S_i is calculated as follows:

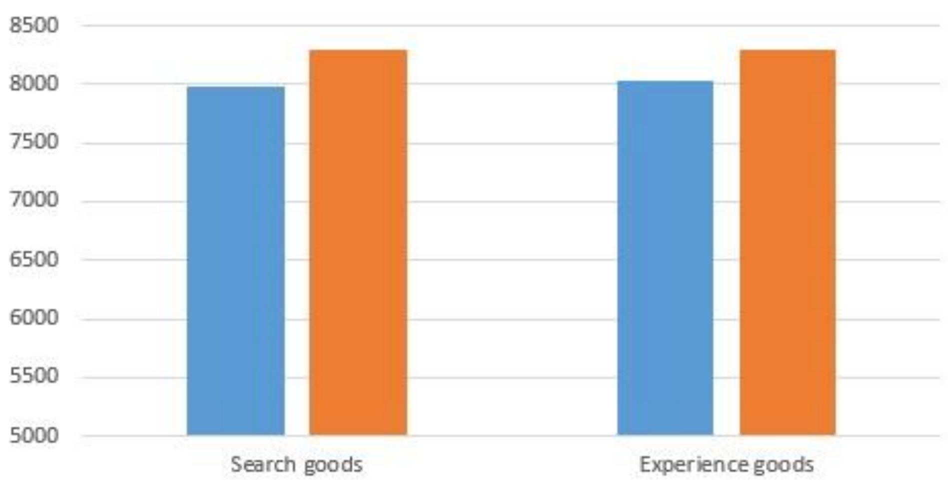
$$NS(S_i) = \max_{j \in S_i}\{u_{ij}\} - \sum_{j \in S_i} c_{ij}$$

In addition, the difference between the net surplus of consumer i with all references and without any references is calculated as follows:

$$\Delta_{NS,i} = NS\left(S_i^* \middle| L = \{L_j, L_k\}\right) - NS\left(S_i^* \middle| L = \{L_j\}\right), \quad j, k \in J$$

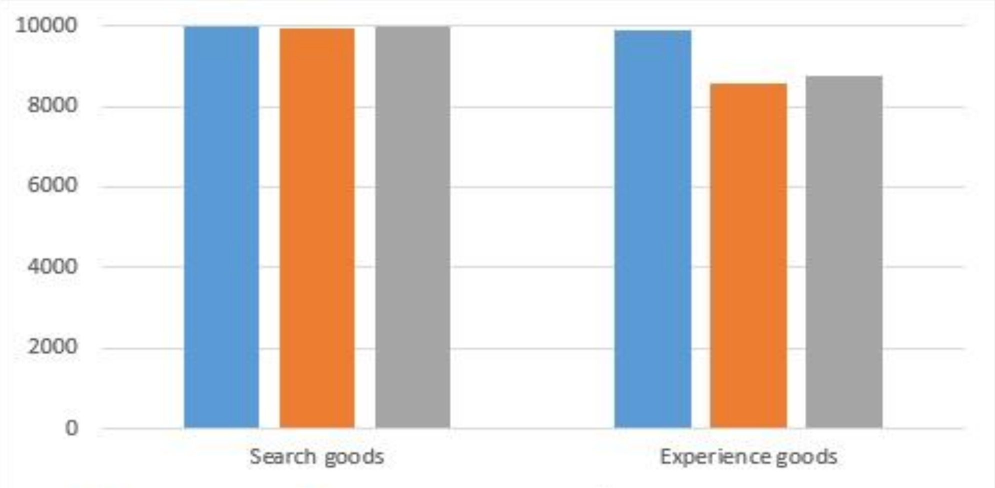
where $(S_i^*|L)$ is an optimal search set given reference L. The first term computes the net surplus of consumer i with all references. The second term computes the net surplus without any references. To investigate and compare the effect of each reference type, I calculated net utility when there were two

references and ruled out each reference for each category. For computing, I generated another 10,000 pseudo-households and made a search set using estimated parameters.



[Figure 5] Effect of the References

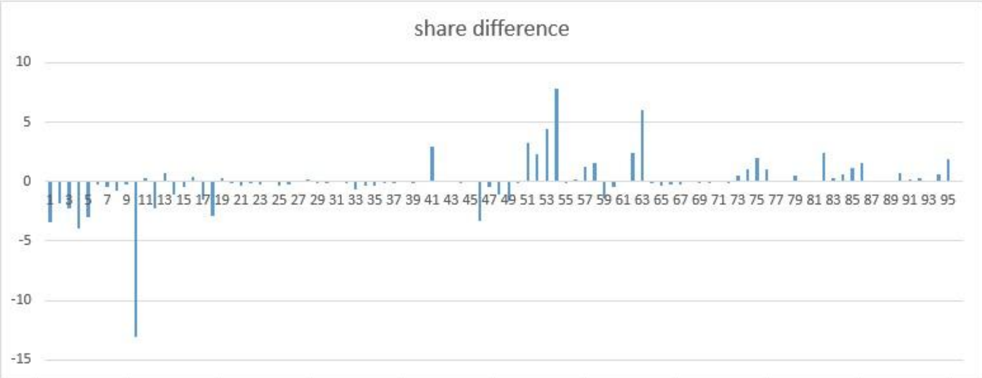
Figure 5 indicates consumers whose net surplus increases if Amazon provides certain references. Overall, consumers' net surplus increases more with references when they search for coffee makers (i.e., search goods). In search goods, 7,975 consumers have an increased surplus when given Amazon.com recommendations and 8,295 consumers have a higher surplus given other consumers' view history list. In experience goods, 8,040 consumers have a higher surplus with consumers' purchase history lists, and 8,292 consumers have an increased net surplus with consumers' view history lists. In both cases, more consumers have an increased surplus when given a view rank list of others who also viewed a certain product. This shows that references based on real search are substantially more helpful to consumers since it has more information about similar products. The net effect of references is significant in both cases; which means that references can substantially help consumers to find the best products with less search cost. In both cases (with and without references), most consumers eventually choose the same product or choose a product with higher utility when two kinds of references are given. To be specific, when they search for coffee makers, 99% of them choose the same product without recommendation from Amazon.com, and 62% choose the same thing regardless of the view-list. Moreover, 0.7% and 20% of them eventually choose a different product with a higher value when given the recommendation and view-list, respectively. In addition, when they search for coffee beans, 35% of them choose the same product regardless of the buy-list information, and 44% choose the same thing regardless of the view-list. Moreover, 41% and 35% of them choose a different product with a higher value when given the buy-list and view list, respectively. This result shows that even though some consumers are misled by references, they still mostly function well as shortcuts. Also, the smaller search volume associated with experience goods results in less availability of information about these products. This means that consumers shopping for experience goods are more likely to be influenced by references when searching for the highest-utility product.



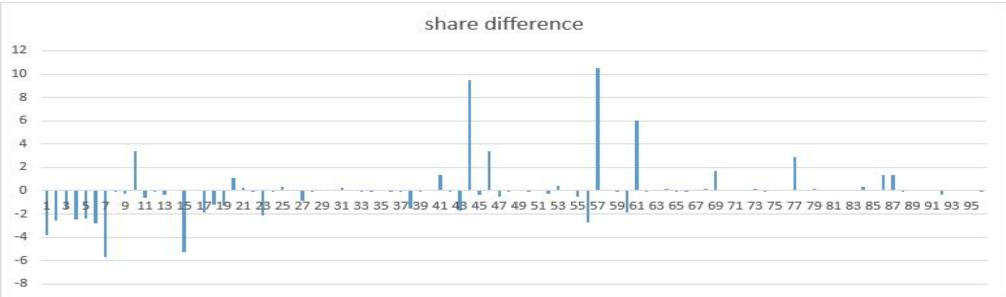
[Figure 6] Number of Consumers Who Decides to Buy

Figure 6 indicates how the number of consumers who buy the product changes due to the existence of shortcuts. References can facilitate purchases by decreasing search cost or helping consumers decide what to buy. I suppose that consumers whose value is higher than zero eventually buy the product for which they have searched. The results show that references increase purchase probability for both coffee makers and coffee beans, while the extent of the increase differs for the two cases. In the case of coffee makers, 9,922 consumers decide to buy without a recommendation from Amazon, and 9,978 consumers are willing to purchase without the search history of other consumers. However, when both references are given on the coffee maker page, the number of consumers who eventually purchase increases to 9,985 (an increase of 0.63% from the condition without an Amazon recommendation and 0.07% from the condition without the view-list). Next, in the case of coffee beans, 8,590 consumers buy when given what other consumers bought with a particular product, and 8,745 consumers buy when given what other consumers viewed with a particular product. However, 9.878 consumers are willing to buy when both types of information are given (an increase of 13.04% from the condition without the buy-list and 11.47% from the condition without the view-list). On account of higher search cost and smaller search set, proper references of experience goods facilitate purchase more. The extent of uncertainty is higher when buy experience goods after reading specification of products on online, record of purchase or browsing is fairly useful.

Finally, I compare the market share of each product under limited and full search conditions. Under the full search condition (i.e., no search cost), consumers search all product before purchase. I assume error term e_{ij} of utility specification to follow normal distribution, so full search condition is same as general probit model.



[Figure 7] Share Difference between Full Search and Limited Search – Search Goods



[Figure 8] Share Difference between Full Search and Limited Search – Experience Goods

Figure 7 and Figure 8 show the share difference of the coffee maker and coffee bean under the full search and limited search conditions. Each graph indicates share difference of each coffee maker and coffee bean under the condition of full search and limited search. Positive value means that the market share increases under full search, and negative value means that the share decreases under the same assumption. Many products show a difference larger than five percent. From the results of the present study, I can conclude that if search costs and limited search conditions are not considered, estimations and predicted results can be biased. In contrast to the results obtained by Kim et al. (2010), the results of the present study do not show a clear relationship between market share and product appearance at reference. For the coffee beans, several products with distinctive positive differences occurred at a very low frequency. However, some coffeemakers with distinctive positive differences occurred at a moderate frequency. Also, many products that seldom appeared at reference showed very small share differences. I suggest three potential reasons to explain this result. First, because the reference is formed based on the consumer preference, the share of products with fewer preferred traits will be low regardless of the search cost or frequency of appearance. Second, unlike Kim et al. (2010), I began this research with the assumption that consumers have limited consideration sets, meaning that products with fewer preferred characteristics will be less likely to be contained within a respective consideration set. Subsequently, a share of these products will be low under full search. Third, some products can be depreciated for unknown reasons. Products that show a large share difference and moderate frequency of appearance should, in theory, appear more frequently at reference. Such products are of high value to consumers, and if more information were given about these products, they would likely be purchased often; however, these products are not presented as often as they should be based on their value.

8. Discussion and Limitation

The present study expanded upon the model by Kim et al. (2010) by adding additional assumptions and analyzing search costs in relation to product characteristics. There are three main contributions of this research to the literature. First, it quantifies search cost and verifies some differences between search goods and experience goods based solely on view-ranking data. This analysis showed that both types of products have different search set sizes due to dissimilarities in search costs between these two types of products. Second, this research took a more life-like approach by assuming a limited consideration set and filtering the decisions of consumers. Based on this approach, it was determined that certified consumer search sets were limited and that the references provided on Amazon.com function well as additional assumptions. Third, this research connects differences in optimal search set size to the effect had by dissimilar changes of references.

The overall results of this study show that experience goods have higher search costs than search goods. This is because it is relatively easy to evaluate the utility of a search good product and compare it to other such products, while experience goods are difficult to judge based on online descriptions alone. Consequently, consumers must take more time and energy to evaluate experience goods before deciding to make a purchase. Additionally, the purchase of experience goods relies strongly on past experience and on individual preferences, while the purchase of search goods relies more on the superiority of the product's specific qualities. In many cases, consumers are less motivated to search for new products, so they instead choose to purchase something they have used before or that is otherwise familiar to them. In sum, high search cost leads to smaller search set size.

When it comes to references—that is, references to similar products given on the page below the product, which are used to reduce search costs—consumer surplus increases in both search goods and in experience goods. However, consumers' choices tend to be affected by experience good references, but not so much by search goods references. This is due to the difference in the optimal search set size between the two types of products. The optimal search set size is larger for search goods than for experience goods, which means that each consumer is better able to assess the maximum utility of a product prior to purchase. Consumers are more influenced by references in the case of experience goods because there is less product information available to them. In the same vein, references given for experience goods lead more consumers to buy the product. To satisfy consumers, references should give accurate information and should help consumers find the most satisfying products, especially in terms of experience goods.

There were some limitations of this research. The first limitation was that, due to insufficient data, I was not able to take repeated searches into account. Consumers often click on links or product pages they have already viewed in order to compare products or evaluate the exact benefit of one product over another. If repetitive searches are concentrated on several products, the utility of these products can be overestimated. The second limitation was that I did not consider consumer attitude toward risk and did not exclude prior consumer knowledge, both of which can affect optimal search set size. Risk-averse consumers search more products to make good decisions, especially if they are going to buy an expensive item. Most consumers also already have some product knowledge from other sources before using Amazon. Therefore, risk-averse attitudes amongst consumers are likely to increase the optimal search set size, and prior consumer knowledge is likely to decrease the optimal search set size. In this research, I successfully compare optimal search set sizes depending on search cost. However, complementing this point will lead to improved results. Measuring the prior knowledge of consumers and creating a model to account for risk-averse attitudes will be important steps in future research on this topic.

References

Baye, M., Morgan, J.. and Scholten, P., "Information, search, and price dispersion", *In T. Hendershott, ed., Handbook on Economics and Information Systems Amsterdam:* Elsevier, 2006.

Berry S., Levinson J. and Pakes A. (1995), "Automobile Prices in Market Equilibrium", *Econometrica*, 63(4): 841-890.

Bronnenberg, Bart J. and Wilfried R. Vanhonacker (1996): "Limited Choice Sets, Local Price Response and Implied Measures of Price Competition," *Journal of Marketing Research*, 33 (May): 163-173.

Chan, Y., T. and Park Y. (2015), "Consumer Search Activities and the Value of Ad Positions in Sponsored Search Advertising", *Marketing Science*, 34(4): 606-623.

Draganska M. and Klapper D. (2011), "Choice Set Heterogeneity and The Role of Advertising: An Analysis with Micro and Macro Data". *Journal of Marketing Research*, 48(4): 653-669.

Goeree M. S. (2008), "Limited Information and Advertising in the U. S.

Personal Computer Industry", *Econometrica*, 76(5): 1017-1074.

Goolsbee A. and Petrin A. (2004), "The Consumer Gains from Direct Broadcast Satellites and the Competition with Cable TV", *Econometrica*, 72(2): 351-381.

Hauser J. R, Urban G. L. and Weinberg B. D. (1983), "Three-dimensional Information Acquisition: An Application to Contraceptive Decision Making", *Advances in Consumer Research*, 10(1): 618-623.

Hauser J. R. and Wernerfelt B. (1990), "An Evaluation Cost Model of Consideration Sets", *Journal of Consumer Research*, 16(4): 393-408.

Hoffman, Donna L. and Thomas P. Novak (1996), "Marketing in Hypermedia Computer-Mediated Environments: Conceptual Foundations", *Journal of Marketing*, 60(July): 50-68.

Howard J. A. and Sheth J. N. (1969), "The Theory of Buyer Behavior", New York: John Wiley.

Huang P., Lurie H. N. and Mitra S. (2009), "Searching for Experience on the Web: An Empirical Examination of Consumer behavior for search and Experience Goods", *Journal of Marketing*, 73(March): 55-69.

Honka, E. and Chintagunta, P., "Simultaneous or Sequential? Search Strategies in the US Auto Insurance Industry”, Working paper.

Kahneman D., “Attention and Effort, Englewood Cliffs”, NJ: Prentice Hall, 1973.

Kim J. B, Albuquerque P. and Bronnenberg, B. J. (2010), “Online Demand under Limited Consumer Search”, *Marketing Science*, 29(6): 1001-1023.

Kim J. B, Albuquerque P. and Bronnenberg, B. J., “The Probit Choice Model under Sequential Search with an Application to Online Retailing”, Working Paper.

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

Leath, A. S. (2005), “Search and Experience Goods: Evidence from the 1960’s and 70’s”, *Journal of Applied Business Research*, 21(1): 45-51.

Mehta N., Rajiv S. and Srinivasan K. (2003), “Price Uncertainty and Consumer Search: A Structural Model of Consideration Set Formation”, *Marketing Science*, 22(1): 58-84.

Moraga-González, J. L., Sándor, Z. and Wildenbeest, M. R. (2013), “Semi-nonparametric Estimation of Consumer Search Costs”. *Journal of Applied*

Economy, 28: 1205–1223.

Moraga-González, J. L., Sándor, Z. and Wildenbeest, M. R., "Consumer Search and Prices in the Automobile Market", Working Paper.

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

Nelson, P. (1974), "Advertising as Information", *Journal of Political Economy*, 82(4): 729-754.

Nevo A. (2001), "Measuring Market Power in the Ready-to-Eat Cereal Industry", *Econometrica*, 69(2): 307-342.

Rust, J. (1987), "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher", *Econometrica*, 55(5): 999-1033.

Roberts, J. H. and Lattin J. M. (1991), "Development and Testing of a Model of Consideration Set Composition", *Journal of Marketing Research*, 28(4): 429-440.

Seiler, S. (2013), "The Impact of Search Costs on Consumer Behavior: A Dynamic Approach", *Quantitative Marketing and Economics*, 11: 155-203.

Smith G. E., Venkatraman M. P. and Dholakia R. R. (1999), "Diagnosing the Search Cost Effect: Wating Time and the Moderating Impact of Prior Category

Knowledge”, *Journal of Economic Psychology*, 20: 285-314.

Song Y. and Carl.F. M., (2011), “A Dynamic Model of Sponsored Search Advertising”, *Marketing Science*, 30(3): 447 – 468.

Weathers. D., Sharma, S. and Wood L. S. (2007), “Effects of online communication practices on consumer perceptions of performance uncertainty for search and experience goods”, *Journal of Retailing*, 83(4): 393-401.

Weitzman M. L(1979), “Optimal Search for the Best Alternative”, *Econometrica*, 47(3): 641-654.

Wildenbeest M. R. (2011), “An empirical model of search with vertically differentiated products”. *The RAND Journal of Economics* 42: 4729-4757.

Wright P. and Barbour F. (1977), “Phased Decision Strategies: SeQuels to Initial Screening,” in Multiple Criteria Decision Making: North Holland TIMS Studies in the Management Science”, M. Starr and M. Zeleny, eds.Amsterdam: North-Holland Publishing Company, 91-109.

국 문 초 록

온라인 상품 구매 시의 최적 검색량과 추천 효과의 차이

- 탐색재와 경험재의 비교를 중심으로

물건 구매 시 소비자들은 인지 능력의 한계로 인해 제한된 수의 상품만을 고려하게 된다. 또한, 탐색 비용(search cost)의 존재로 인해 합리적인 소비자들은 제한된 수의 상품만을 탐색하게 된다. 탐색 비용은 크게 외부적 비용과 내부적 비용으로 나눌 수 있는데, 여기에는 상품을 찾는 행위 자체에 들어간 비용뿐만 아니라 상품과 관련된 정보를 처리하고 판단하는 비용 또한 포함된다.

본 연구는 Kim et al. (2010)의 제한된 탐색(limited search)을 설명하는 구조적 모형에 제한된 고려 상품군(limited awareness set)이라는 새로운 가정을 추가하여 탐색재(search goods)와 경험재(experience goods)의 탐색 비용을 추정하였다. 경험재의 경우 상품과 관련된 정보

를 처리하는 데에 더 많은 시간이 걸린다는 Huang et al. (2009)의 연구와 일관성 있게 경험재의 탐색 비용이 탐색재에 비해 유의하게 높았다. 추정된 모수로 시뮬레이션 및 역사실적 실험(counterfactual experiment)을 진행한 결과, 경험재의 경우 탐색 비용이 높은 만큼 최적 검색 상품의 개수(optimal search set size)가 탐색재에 비해 적었고, 그만큼 아마존(Amazon.com)에서 자체 제공하는 참조 상품(references)이 소비자들의 상품 선택에 많은 영향을 미치고 소비자들의 구매를 촉진시켰다. 이를 통해 정확한 상품 추천이 경험재의 경우에 더 중요하다는 결론을 도출하였다.

주요어: 탐색재, 경험재, 탐색 비용, 구조적 모형, 참조 상품의 영향, 임의 계수 이산 선택 모형

학번: 2014-20421