# Online Consumer Search Depth: Theories and New Findings

JIE "JENNIFER" ZHANG, XIAO FANG, AND OLIVIA R. LIU SHENG

JIE "JENNIFER" ZHANG is an Assistant Professor of Information, Operation and Technology Management in the College of Business Administration, University of Toledo. She received her Ph.D. in Computer Information Systems from the University of Rochester in 2003. She employs analytical and empirical techniques to examine a number of issues in e-commerce and management of information systems. Her research appears in *Journal of Economics and Management Strategies, Journal of Management Information Systems, Information Resources Management Journal,* and *Journal of Computer Information Systems.*

XIAO FANG is an Assistant Professor of Information Systems at the College of Business Administration, University of Toledo. He received his Ph.D. in Management Information Systems from the University of Arizona in 2003. His current research interests are in the areas of business intelligence and knowledge management. He has published in journals such as *INFORMS Journal on Computing, ACM Transactions on Internet Technology, Communications of the ACM,* and *Journal of the American Society for Information Science and Technology.*

OLIVIA R. LIU SHENG is Emma Eccles Jones Presidential Chair of Information Systems at the David Eccles School of Business, University of Utah. She received her Ph.D. in Computers and Information Systems from the University of Rochester. Her research focuses on global information technology and management, including distributed enterprise data and knowledge management, e-commerce, telemedicine, and distributed learning. She has published over 40 journal articles and has received funding from the U.S. Army, NSF, IBM, Toshiba, Sun Microsystems, Hong Kong Research Grants Council, Asia Productivity Organization, and SAP University Alliance. Dr. Sheng is an Associate Editor for *IEEE Transactions on Engineering Management, Electronic Commerce Research and Applications, Electronic Commerce Management Research, Telecommunication Systems,* and *Journal of Information Technology and Management.* She has guest or co-guest edited a number of special issues of *IEEE Transactions on Knowledge and Data Engineering, Decision Support Systems, Journal of Parallel and Distributed Computing,* and *Telecommunication Systems* on multimedia information systems and health-care decision support.

ABSTRACT: The continuous growth of e-commerce makes it critical for firms to understand consumers' search behavior so that e-commerce Web sites and the underlying information systems can be designed to better cater to consumers' needs. This paper extends the classic search model to analyze online consumer search behavior. The analytical results suggest how consumers' search depth is influenced by a variety of factors such as search cost, individual consumer difference, and product characteristics. Evidence is provided using clickstream data of online searches and purchases

*Journal of Management Information Systems* / Winter 2006–7, Vol. 23, No. 3, pp. 71–95.

0742–1222 / 2007 $9.50 + 0.00.
DOI 10.2753/MIS0742-1222230304

of music CDs, computer hardware, and airline tickets during the period from July 2002 to December 2002 collected by an Internet marketing company, ComScore Inc. Compared with the search depth reported in previous works, this study finds that consumers are searching more intensely before purchasing online. This reflects the evolution of Internet users and the growth of online retail business.

KEY WORDS AND PHRASES: clickstream data, consumer search behavior, online search behavior, search depth, search model.

SINCE ITS BEGINNING IN 1995, E-COMMERCE has experienced a steady increase in the United States and is projected to continue growing at double-digit rates. A recent report issued by the Department of Commerce Census Bureau indicates that online retail sales in the first quarter of 2004 were $15.5 billion, up 27.2 percent from the first quarter of 2003. The rapidly growing online retail sales bring enormous opportunities and challenges to both consumers and online retailers. As more and more online businesses emerge, consumers have more choices of Web stores to shop. Yet, in order to make a decision, a consumer has to collect and digest a vast amount of information on the detailed offering of each individual site. Therefore, consumers who have an item in mind to purchase have to engage in a great deal of search in order to find a site that offers an appropriate price and satisfactory product or service quality (e.g., shipping service, product availability, and retailer reputation). In the meantime, because search is the first step in a consumer's shopping experience and competitors are just "one click away," it is important for online retailers to understand consumers' search behavior and make strategic decisions accordingly. Understanding consumers' search behavior is critical for online retailers to design e-commerce Web sites and their underlying information systems to better cater to consumers' needs. Therefore, a critical problem for both researchers and practitioners is to study how consumers search, and especially what determines search depth—that is, the number of sites sought before a purchase.

Existing economic theory [22, 27] modeled consumers' search behavior as a compromise of the anticipated utility gain through price reduction and the additional search cost. Those models assumed that consumers are only searching for a single attribute (e.g., price). However, recent empirical findings by Smith and Brynjolfsson [23] reveal that consumers are not choosing the online retailer that offers the lowest price, but actually balancing price and quality factors in making their purchase decisions. This paper builds a richer model of consumers' online search incorporating both price and quality attributes to analyze consumers' search behavior.

A wealth of marketing literatures have studied consumers' search in the traditional offline market (see Newman [20] for a review of selected research). A consistent empirical finding of those works is that consumers exhibit very limited search activity in the offline market [4, 20]. Moorthy et al. [18] explained this low search level based on distributions of consumers' prior brand perception. Numerous other works have studied the effects of various factors on consumers' search of traditional stores, including price, perceived risk, past experiences, perceived search benefits, level of education, search cost, and so on (see a good summary in Beatty and Smith [4]). This paper empirically examines variables related to consumers' online search and compares the results with the previous findings in the offline world.

There has been consensus that the Internet has greatly reduced consumers' search costs [2, 3], which raises the question about whether decreased search costs on the Internet increase consumers' search. In contrast to the vast literature on consumer search in the offline world, only a few investigations have been conducted on consumers' online search behavior. Bhatnagar and Ghose [5] reported 63 percent of the respondents searched the Web sites in the same category less than twice per month. Ratchford et al. [21] found that those who use the Internet to search for automobiles are younger and more educated, and search more in general. However, the analysis also indicated that those consumers would have searched even more without the presence of the Internet.

Those previous studies on searches employed methodologies such as surveys, experiments, and interviews to collect data. However, due to various limitations, those self-reported data have been found to be poor measures of actual search activities [20]. In the online market, consumers' Web navigating and purchasing behavior can be easily and accurately tracked using server-side or client-side programs. These tracked data—namely, clickstream data—provide micro-level information to study consumers' online shopping behavior [8]. Using a set of clickstream data collected by Media Matrix Inc. (now part of ComScore Networks) during the period from July 1997 to June 1998, Johnson et al. [14] studied the search depth of consumer online purchases of books, CDs, and air travel services. Surprisingly, they found that consumers on average only searched less than two stores for the goods in the three categories (1.2, 1.3, and 1.8 for books, CDs, and airline tickets, respectively). They explained the result with customer loyalty, immaturity of the online population, and consumers' reluctance to search.

Compared with Johnson et al. [14], whose focus was on the very beginning stage of e-commerce, this paper looks at a relative mature stage of e-commerce after an explosive growth period and the dot-com shakeout in spring 2000. Using the clickstream data collected by the same company during a later time period, this paper finds that consumers on average searched 2.1 stores for a purchase of CDs and 3.3 stores for both airline tickets and computer hardware products, almost double the results found in Johnson et al. [14]. The difference between the two signifies the maturity of Internet users, the fast growth of e-commerce, and the adoption of new Internet technologies.

## New Findings from the More Recent Clickstream Data

THIS RESEARCH EMPLOYS A DATABASE COLLECTED by the Internet marketing company ComScore Media Matrix (www.comscore.com), which records consumers' Web navigating and purchasing logs tracked using client-side programs installed on the recruited households' home computers. The ComScore database used in this research captures 100 million Web site visits and 342,706 transactions conducted by 100,000 households across the United States during a six-month period from July to December 2002 and demographical data of these households. It is more current and richer in content than the database used by Johnson et al. [14], which captures clickstream data of only 10,000 households from July 1997 to June 1998. Compared with the survey data used in a majority of the previous studies [4, 5, 18], clickstream data has the advantage of providing a larger and more objective sample of data about consumers' actual online shopping behavior. Moreover, the availability of household demographic data (such as household size, income, highest education, oldest age, Internet connection speed, and presence of children) enables us to capture the heterogeneity of the individual search behavior in a more direct way than that adopted by Johnson et al. [14].

Specifically, this study focuses on three categories of goods—music CDs, computer hardware, and air travel. For each product category, the corresponding retailer Web sites were chosen according to (1) leading retailer Web sites in this category listed by BizRate.com, (2) retailer Web sites in this category used by previous research, and (3) retailer Web sites that had sales records in this category in our data set. In the ComScore data (i.e., panel data used in this research), consumer Web site visiting behaviors are recorded at session level, not at click level, thus we cannot judge the purpose of a visit to a general-purpose retailer. To avoid inflating the number of sites visited during a search session and to make our comparison with previous studies on search depth robust, the Web site list is narrowed to those of specialized retailers in each category and we remove those general-purpose retailer Web sites such as Yahoo stores, Walmart.com, and eBay.com. As shown in Tables 1 and 2, after removing those general-purpose retailers, the search depths for music, computer hardware, and air travel stores are reduced by 8.7 percent, 5.7 percent, and 5.7 percent, respectively.

We finalize with 16 music retailer Web sites, 24 computer hardware retailer Web sites, and 29 air travel retailer Web sites. Table 3 lists the Web sites with the number of visits by each site during the research periods.

To examine consumers' search behavior, we need to define search sessions. Technically, a session is defined as a sequence of consecutive Web accesses by the same visitor. Computer scientists have been using the widely applied rule of thumb [12, 25] to identify sessions that stipulates that the maximal session length cannot exceed 30 minutes. However, as pointed out by Johnson et al. [14], this technical definition of session is too narrow to analyze search behavior, as a shopper's decision period may span multiple days, weeks, and so on. Johnson et al. [14] defined a search session as a series of store visits over a span of days that eventually leads to a purchase. Specifically, they defined a search session covering one calendar month. In this paper, we adopt this monthly level search session definition because (1) we used the database collected by the same company as Johnson's study, and (2) the key justification of using the monthly level search session definition in Johnson's study—less than 1 percent of monthlong sessions containing more than one transaction—still holds in the 2002 ComScore database used in our study.

Table 1. Statistics of Visits to General-Purpose Retailers

| Product category | Number of removed visits to general-purpose retailer | Number of visits used in this study | Total number of visits | Percentage of visits removed |
|---|---|---|---|---|
| Music | 3,542 | 29,163 | 32,705 | 10.8 |
| Computer hardware | 1,583 | 15,176 | 16,759 | 9.4 |
| Air travel | 5,085 | 60,077 | 65,162 | 7.8 |

Table 2. Search Depth Reduced Due to Removing General-Purpose Retailers

| Product category | Search depth (with general-purpose retailers) | Search depth (without general-purpose retailers) | Reduction percentage |
|---|---|---|---|
| Music | 2.3 | 2.1 | 8.7 |
| Computer hardware | 3.5 | 3.3 | 5.7 |
| Air travel | 3.5 | 3.3 | 5.7 |

We define search depth as the number of *unique* retailer Web sites within a product category visited during a search session. By analyzing the 2002 ComScore database, we find that even though we undercount the search depth by removing general merchandisers from the retailer lists, consumers still search more intensively than they did at the initial period of e-commerce. The average search depths for music, computer hardware, and air travel Web sites are 2.1, 3.3, and 3.3, respectively (Figure 1), which increase dramatically from the search depth results reported in Johnson et al. [14]—the average search depth for music increases by 62 percent and the average search depth for air travel almost doubles.

As consumers search more stores, we also find their loyalty to one particular store is compromised. We observe only 37 percent of the music CD shoppers, 7 percent of the computer hardware shoppers, and 19 percent of the air travel ticket shoppers remain loyal to one retailer Web site throughout the research period, as compared with 70 percent of the music CD shoppers and 42 percent of the air travel shoppers in Johnson et al. [14].

We now build a formal research model to study consumers' online searching behavior. The different results found above from the previous studies can be explained by theories derived from that model in terms of the growth of e-commerce, technology advancement, and the maturity of the Internet users: (1) Johnson et al. [14] studied consumers' search behavior at the starting stage of e-commerce before the Internet boom in late 1998, while our research reexamines the online search problem when e-commerce had overcome the boom-and-bust cycle. Over a mere five years, the Internet

Table 3. Retailer Web Sites and Their Corresponding Number of Visits

| Web site | Number of visits |
|---|---|
| Music | |
| Amazon.com | 16,042 |
| Columbiahouse.com | 3,942 |
| CDnow.com | 3,023 |
| BMGmusic.com | 1,504 |
| Bestbuy.com | 1,362 |
| MP3.com | 1,260 |
| Buy.com | 678 |
| CDuniverse.com | 560 |
| BN.com | 301 |
| Fye.com | 203 |
| Samgoody.com | 190 |
| Mymusic.com | 36 |
| CDnowpbc.com | 35 |
| Hmv.co.uk | 16 |
| CDEurope.com | 8 |
| Musicblvd.com | 3 |
| Computer hardware | |
| Dell.com | 5,641 |
| HP.com | 1,607 |
| Bestbuy.com | 1,341 |
| Officedepot.com | 1,001 |
| Apple.com | 796 |
| Tigerdirect.com | 728 |
| Circuitcity.com | 646 |
| Sears.com | 643 |
| Compaq.com | 422 |
| CompUSA.com | 414 |
| Sonystyle.com | 404 |
| Compgeeks.com | 220 |
| IBM.com | 201 |
| Outpost.com | 192 |
| Vikingop.com | 192 |
| Computers4sure.com | 153 |
| Digikey.com | 137 |
| Epson.com | 137 |
| PCconnection.com | 83 |
| Mwave.com | 73 |
| Warehouse.com | 73 |
| Insight.com | 32 |
| Accessmicro.com | 29 |
| Futureshop.ca | 11 |
| Air travel | |
| Expedia.com | 13,909 |
| Orbitz.com | 10,114 |
| Southwest.com | 5,833 |
| Hotwire.com | 4,504 |

| Web site | Number of visits |
|---|---|
| Delta.com | 3,166 |
| Travelzoo.com | 3,102 |
| Priceline.com | 3,045 |
| AA.com | 2,820 |
| ITN.net | 2,018 |
| Alaskaair.com | 1,531 |
| Continental.com | 1,382 |
| Travelnow.com | 1,322 |
| Jetblueairways.com | 1,048 |
| Americawest.com | 972 |
| Trip.com | 810 |
| Onetravel.com | 796 |
| United.com | 773 |
| NWA.com | 754 |
| Spiritair.com | 707 |
| ATA.com | 703 |
| Lowestfare.com | 245 |
| Britishairways.com | 180 |
| Site59.com | 132 |
| Aavacations.com | 90 |
| Aircanada.ca | 70 |
| Bookryanair.com | 51 |
| Easyjet.com | 27 |
| City.net | 17 |
| Previewtravel.com | 6 |

exploded—the number of Internet hosts expanded from 19 million in July 1997 to 162 million in July 2002, according to a survey conducted by the Internet Systems Consortium Inc. With the development of the Internet, Web sites have become more user-friendly and easier for consumers to navigate. (2) Broadband Internet was more widely adopted during the years between the five-year gap of the two studies. By 2003, broadband penetration had reached 13 percent—a growth of about 150 times from the end of 1997. Hence, connection speed has dramatically increased due to broadband adoption, making searching the Web much faster. (3) The rapid development of Internet technology has penetrated more households. Users are more familiar with the Internet environment and are more proficient in searching.

## Consumer Online Sequential Search Model

In the online environment, the cost of becoming an online retailer is relatively low compared with building a brick-and-mortar store in the physical world. This low entry cost increases the number of online retail stores with various levels of service quality and reputation, which complicates consumers' search activities. Consumers will not choose a site with the lowest price but unacceptable service quality; they would rather choose a site with better service and reputation for a higher price [23].

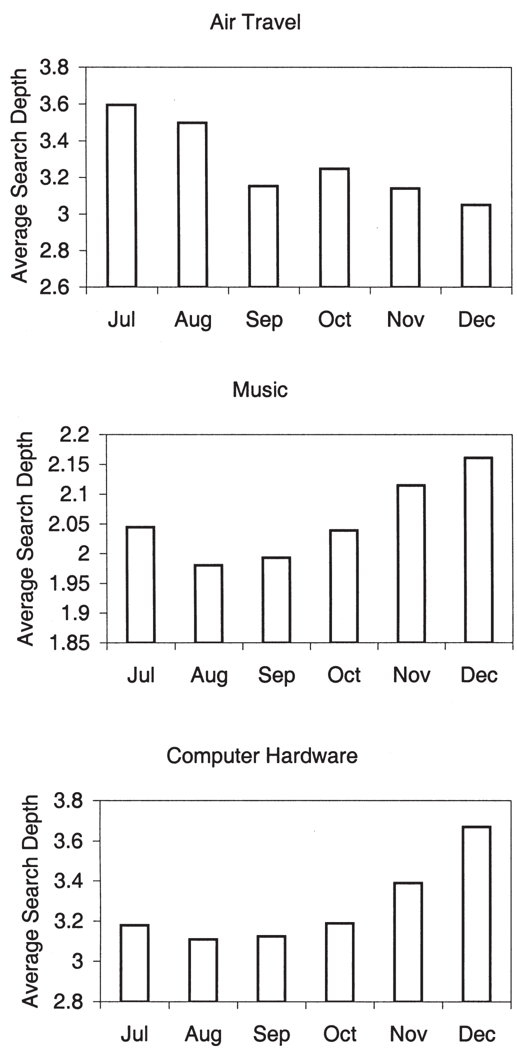


*Figure 1.* Average Number of Online Stores Visited During Each Observed Shopping Month

For example, the few top online retailers that sold the most in each of the three categories are all recommended and rated above average by the ConsumerReports.org e-ratings, which evaluates the overall quality of a Web site based on "credibility, usability, content and how these components come together to create a satisfying, efficient and effective online experience."

Building on the traditional search models [22, 27], we develop an analytical model to characterize consumer search and purchase decisions incorporating both price and nonprice attributes. Nonprice attributes include retailer reputation, brand, service quality, security, and privacy protection of the Web sites.

Suppose that there is a large enough number of online retailers and each one of them has quality $q$ and charges price $p$, where either $p$ or $q$ is an independent and identically distributed random variable. The two variables $p$ and $q$ have a joint distribution function $F(p, q)$ and a positive correlation $\rho > 0$. That is, a better-known or higher-quality retailer usually charges a price premium. A consumer's utility of purchasing the product from a store that offers the good with price $p$ and quality $q$ is

$$u(p,q) = v - p + \theta q, \tag{1}$$

where $v$ and $\theta$ denote the consumer's value and quality preference, respectively. The valuation and the preference parameters vary with different consumers and denote their types. For example, other things being equal, high $v$ consumers are willing to pay more for the searched item, and high $\theta$ consumers care more and are willing to pay more for quality than low $\theta$ ones. Let

$$m(p,q) = p - \theta q, \tag{2}$$

where $m$ is the measure a consumer uses to compare the stores by considering both price and nonprice attributes. Consumers will search the stores and try to purchase at a store with the lowest quality-adjusted price $m$. Given the joint distribution function of $p$ and $q$ $F(p, q)$, suppose we can obtain a cumulative distribution function of $m$—say, $F_M(\cdot)$ and a density function $f_M(\cdot)$.

At the early stage of the e-commerce era, online retailers were worried that the fierce price competition would squeeze out all the profit, because the Internet brings competitors so close that they are simply "one click away." Yet it turns out not to be true. Empirical studies [6, 7, 9, 24] have demonstrated significant price dispersion among online retailers. The existence of search cost has been ascribed as a main reason for this phenomenon. Because it takes time and effort to locate and examine each online retailer, there is nonnegligible search cost involved. We model the cost of searching an online store as a constant $c$.

Suppose a consumer's best offer after searching $n$ Web stores is

$$\left(\widehat{p,q}\right) = \underset{(p,q)}{\arg\min}\left\{m(p_1,q_1), m(p_2,q_2)\ldots m(p_n,q_n)\right\};$$

the consumer will receive net utility $v - m(\widehat{p,q})$ if he or she purchases from that site and stops searching. Or the consumer can continue searching another online store at a cost $c$. In this case, the consumer may accept a better offer $(p', q')$, which has $m(p', q') < m(\widehat{p,q})$. Otherwise, the consumer has to go back to purchase from the previous best offer (in the online environment, there is negligible cost for the consumer to go back to any of the previous Web sites searched before). We express the value function of this complete search and purchase process as the following Bellman equation:

$$V\left(\widehat{p,q}\right) = \max\left\{v - m\left(\widehat{p,q}\right), \int_{m(p',q')<m\left(\widehat{p,q}\right)} V\left(p',q'\right) dF\left(p',q'\right) \right.$$
$$\left. + \int_{m(p',q')\geq\left(\widehat{p,q}\right)} V\left(\widehat{p,q}\right) dF\left(p',q'\right) - c\right\}. \tag{3}$$

A consumer with a best (price, quality) offer $(\widehat{p,q})$ from the previous $n$ searches will compare his or her current maximal utility $v - m(\widehat{p,q})$ with the expected gain from searching one more store to decide whether it is worthwhile to continue searching. At a cost $c$ for one more search, the consumer may obtain a better offer $(p', q')$ with probability $\int_{m(p',q')<m(\widehat{p,q})} dF(p',q')$ or a worse one with probability $\int_{m(p',q')\geq m(\widehat{p,q})} dF(p',q')$, in which case he or she will give up and return to the previous best offer $(\widehat{p,q})$. The consumer will trade off this expected offer from the next search with the search cost. The solution to this dynamic programming problem should make consumers indifferent to the current offer and the expected gain from one more search. We call this solution as stopping price and denote it as $m_r$. That is,

$$v - m_r = \int_{m\geq m_r} \left(v - m_r\right) dF_M\left(m\right) + \int_{m<m_r} \left(v - m\right) dF_M\left(m\right) - c. \tag{4}$$

A consumer keeps searching the stores sequentially until he or she finds a store offering quality-adjusted price $m(p,q)$ smaller than or equal to $m_r$. This stopping rule is consumers' optimal strategy proved by Weitzman [27].

Suppose the support for $m$ is $[l, u]$. Simplifying Equation (4), we get

$$\int_l^{m_r} \left(m_r - m\right) dF_M\left(m\right) = c. \tag{5}$$

That is, a consumer will continue searching stores until his or her expected utility gain from searching one more store (the expected utility improvement from better quality and price combination offered by the next store) equals the marginal cost $c$. By simplifying Equation (5) further, we can solve the value of $m_r$ from

$$\int_l^{m_r} F_M\left(m\right) dm = c. \tag{5$'$}$$

Because we cannot observe this stopping price in our data, we go beyond the traditional search model and further derive the search depth that is available from our database. Given the optimal search strategy specified above, the number of stores searched by a consumer before purchasing from a satisfactory one follows a geometric distribution. The expected number of stores searched $N$ is derived as follows:

$$E\left(N\right) = \sum_{i=1}^{\infty} i\left(1 - F_M\left(m_r\right)\right)^{i-1} F_M\left(m_r\right) = \frac{1}{F_M\left(m_r\right)}. \tag{6}$$

Because $(\partial E(N)/\partial m_r) = -(f_M(m_r)/F_M(m_r)^2) < 0$, Equation (6) suggests that as the stopping value $m_r$ increases, the consumer is expected to search fewer stores. In the extreme case, if a consumer only purchases at a store that offers the lowest quality-adjusted price among all the retailers, that is, his or her stopping value is the lower bound of the support of $m$, he or she is expected to search infinitely: $E(N) \rightarrow \infty$.

This stopping utility value $m_r$ and the expected search number $E(N)$ cannot be obtained analytically without specifying a distribution function $F(p, q)$. Hence, we derive the properties of this expected search depth $E(N)$ indirectly and propose the following findings.

> *Proposition 1: A consumer's search depth increases with his or her quality preference $\theta$, and decreases with his or her marginal search cost* c.

Not only do consumers' inherent characteristics and marginal search cost influence consumers' search depth, Equation (5) implies that the distribution function $F_M(m)$ also affects the value of the stopping utility $m_r$, and further affects the search depth. Under some restrictions on the distribution function $F(p, q)$, Propositions 2 and 3 suggest a positive influence of the mean and variance of the quality-adjusted price $m$ on the expected search depth $E(N)$.

> *Proposition 2: If distribution function* $\mathrm{F}_{\mathrm{M}}^{1}(\mathrm{m})$ *first-order stochastic dominates* $\mathrm{F}_{\mathrm{M}}^{2}(\mathrm{m})$*, then the consumers search more stores with the higher expected mean quality-adjusted price* m, *that is,* $\mathrm{E}_1(\mathrm{N}) \geq \mathrm{E}_2(\mathrm{N})$.

For high-end goods, everything else being equal, consumers are willing to spend more time to search for a better price to realize a larger saving or for a higher-quality Web site to mitigate the potential risks involved in the transaction and delivery. Hence searching for an item of a higher expected quality-adjusted price provides higher rewards and stimulates the consumers to search more.

> *Proposition 3: If distribution function* $\mathrm{F}_{\mathrm{M}}^{1}(\mathrm{m})$ *is a mean-preserving spread of the distribution function* $\mathrm{F}_{\mathrm{M}}^{2}(\mathrm{m})$ *and* $\sigma_1^2 \leq \sigma_2^2$*, then consumers are expected to search more stores when there is a higher utility uncertainty, that is,* $\mathrm{E}_1(\mathrm{N}) \leq \mathrm{E}_2(\mathrm{N})$.

Proposition 3 suggests that higher dispersion of the quality-adjusted price increases consumers' search activity. Because $\sigma_m^2 = \theta^2\sigma_q^2 + \sigma_p^2 - 2\theta\rho\sigma_p\sigma_q$, the higher the price and quality correlation $\rho$, the lower the variance of the variable $m$. Therefore, keeping the mean of $m(p,q) = p - \theta q$ constant, a high correlation between price and quality will reduce the variance of quality given the price or vice versa, which saves consumers some search to collection information on both variables. We summarize it into the following corollary:

> *Corollary 1: Other things being equal, a high price–quality correlation will reduce consumers' expected search depth.*

## Research Framework and Hypotheses

BASED ON THE ABOVE PROPOSITIONS as well as existing literature, we derived hypotheses about the factors from our ComScore database that affect consumer online search depth. Because most of the variables in the analytical model are not directly observable in our database, we generated our hypotheses using variables from our database that have a direct relationship with the variables in the analytical model. For example, we cannot measure consumers' search cost directly, therefore we use high-speed Internet adoption (less time in searching) and education level (more skillful in searching), which are measurable and available in our database, to approximate search cost in the hypotheses. Relationships such as this between the empirical observable variables and the variables in the theoretical model are illustrated in Table 4 and then we posit hypotheses based on the connections.

From the preceding analysis, we conclude that the following factors have an effect on search depth.

1. Search cost of visiting each store—variable $c$ in the analytical model. According to Proposition 1, a lower cost of search increases search depth. It is consistent with the empirical findings in previous studies [16, 26]. Broadband Internet connection is unique for online search and one of the important measures of consumer online search cost. Adopting broadband Internet connections or using software agents to facilitate searches are expected to reduce the cost of searching Web stores; therefore, consumers with high-speed Internet connections are expected to search more online stores. Because our data set records the online browsing and shopping behavior of each household, we suggest the following:

*Hypothesis 1: Households with high-speed Internet connection search in greater depth than those households without.*

2. Product characteristics. Under some restrictions, Proposition 2 suggests that the higher expected quality-adjusted price of a product increases the perceived benefit from the search, and therefore increases the search depth. Empirically we use the price paid as a proxy of the value of the goods. Hence, we posit:

*Hypothesis 2: The higher the price a consumer will pay for the search item, the more searches the consumer will conduct.*

3. Consumer characteristics. Some consumers may be more skillful at searching the Internet for useful information or be more willing to spend time surfing than others. Those discrepancies stem from the heterogeneity of each individual consumer's experience with the Internet, value of time, preference for quality, patience, and so on.

In this sense, search depth of the last session is a good measure of a consumer's search propensity, and is expected to be positively correlated with search depth in the current session. As such, we have the following hypothesis:

*Hypothesis 3: Households that searched in greater depth during the last session tend to search in greater depth during this session.*

High-income consumers have higher value for their time. Supposing no strong correlation between search time and income level as demonstrated in our data sample, then high-income consumers incur greater cost in searching, which reduces their search

Table 4. Connection Between the Theoretical Model and the Empirical Hypotheses

| | Variables in the empirical model | Justification from the theoretical model | Hypothesized correlation with search depth |
|---|---|---|---|
| H1 | Broadband Internet adoption | Lower search cost (P1) | Positive |
| H2 | Price | Higher quality–adjusted price (P2) | Positive |
| H3 | Lag of search depth | Higher search skills<br>Higher quality preference (P1) | Positive |
| H4 | Household income | Higher quality preference<br>Higher search cost (P1) | Positive |
| H5 | Household education | Lower search cost (P1) | Positive |
| H6 | Household age | Higher quality preference<br>Higher search cost (P1) | Positive |

intensity by the first part of Proposition 1. However, their high purchasing powers imply that they are less sensitive to price and put a heavier weight on quality when evaluating the choices of online retailers. By the second part of Proposition 1, they will search more. Hence, the influence of income on search depth depends on the net of the cost effect and the quality effect.

*Hypothesis 4: Higher income households search more sites when shopping online.*

Mixed results have been found about the correlation of education with the searches of offline stores [15, 16]. In the online market, more educated persons were innovators or early adopters in the usage of the Internet [13]. They are more skillful at searching the Web for useful information with less time (and cost). Consequently, one would expect education to have a statistically significant effect on Web search behavior.

*Hypothesis 5: Higher education households search more stores when shopping online.*

Older consumers typically have less time and energy to conduct a trip to a physical store, which imposes higher search cost for them in the offline market. This is supported by the findings that age has an inverse relationship with degree of search of traditional retail stores [11, 15, 16, 26]. In the online market, search can be conducted without trips and therefore search cost does not vary significantly for consumers of different ages. Can this inverse relationship between age and search depth in the traditional market still hold true for online search? This question motivates our next hypothesis.

*Hypothesis 6: Older households search more stores than younger ones when shopping online.*

These hypotheses are verified with our consumer clickstream data in the music CD, computer hardware, and air ticket categories. The results are analyzed with the above theories.

## Econometric Specification and Estimates

COMBINING THE SEARCH DEPTH DATA, transaction data, and household demographic data, we obtain our final data sample for studying the factors affecting consumers' search behavior. Table 5 presents the descriptive summary of the variables used in the empirical study.

There are 2,894, 2,878, and 5,919 households that made at least one purchase in the music, computer hardware, and air travel categories, respectively, during the study period. The mean amount paid per purchase denotes the value of the goods, and it increases from the music category to the computer hardware to the air travel category, with the values $98.57, $398.53, and $518.47, respectively. On average, households that purchased music products are younger, less educated, have lower income, and are less likely to have a broadband connection than those who purchased in the other two categories.

Comparing the results across product categories, search depth increases in the order of music category, computer category, and air travel category, which can be explained by the price distribution and product characteristics of the three categories. Music CDs have the lowest mean price paid, as well as a low variance. Therefore, the expected gain from price reduction through one additional search is also low. In contrast, computer hardware products and airline tickets have higher-quality uncertainty: computer hardware products are more differentiated across brands, air travel tickets are highly differentiated in terms of departure and arrival time, number of connections, and so on [10]. An additional search has more potential to increase consumers' utility through price reduction and quality improvement. By Propositions 2 and 3, in some cases, high mean and variance of quality-adjusted prices give rise to more active search, which is supported by the finding that computer hardware products and air travel categories have much greater search depth.

By replacing the search propensity variable in Johnson et al. [14] with the search depth in the last session and adding such controlling variables as price paid and households demographic variables, we obtain the following regression model:

$$\begin{aligned} N_{ij} = \beta_0 + \beta_1 \ln p_{ij} + \beta_2 N_{i,j-1} + \beta_3 Connection\_Speed_i + \beta_4 Household\_Income_i \\ + \beta_5 Household\_Education_i + \beta_6 Household\_Oldest\_Age_i \\ + \beta_7 Household\_Size_i + \beta_8 Child\_Present_i. \end{aligned} \tag{10}$$

We tested model (10) to verify the hypotheses and provide the results in Table 6. We used the natural logarithm of the purchased item price to capture the nonlinear relationship of price paid and search depth. Because about one-third of the households did not report the education information, we ran the linear regression model (10) for

Table 5. Descriptive statistics for the three product categories

Panel A. Descriptive statistics for music CD category

| Variables | Mean | Number | Standard deviation | Maximum | Minimum |
|---|---|---|---|---|---|
| $N_{ij}$ | 2.06 | 3,851 | 1.22 | 10 | 1 |
| $\log(p_{ij})$ | 3.62 | 3,851 | 1.31 | 8.45 | 0 |
| $N_{i,j-1}$ | 0.57 | 3,851 | 1.19 | 10 | 0 |
| *Household_Income* | 4.33 | 3,851 | 1.63 | 7 | 1 |
| *Connection_Speed* | 0.45 | 3,851 | 0.50 | 1 | 0 |
| *Household_Education* | 2.84 | 2,695 | 1.39 | 5 | 0 |
| *Household_Oldest_Age* | 6.48 | 3,851 | 2.50 | 11 | 1 |
| *Child_Present* | 0.45 | 3,851 | 0.50 | 1 | 0 |
| *Household_Size* | 3.03 | 3,851 | 1.36 | 6 | 1 |

Correlation matrix for computer CD category

| | $N_{ij}$ | $\log(p_{ij})$ | $N_{i,j-1}$ | *Household_ Income* | *Household_ Education* | *Connection_ Speed* | *Child_ Present* | *Household_ Size* | *Household_ Oldest_Age* |
|---|---|---|---|---|---|---|---|---|---|
| $N_{ij}$ | 1.000 | 0.005 | 0.306 | –0.031 | 0.027 | 0.058 | 0.046 | 0.044 | 0.003 |
| $\log(p_{ij})$ | 0.005 | 1.000 | 0.014 | 0.073 | –0.043 | 0.010 | –0.044 | –0.043 | 0.022 |
| $N_{i,j-1}$ | 0.306 | 0.014 | 1.000 | –0.016 | –0.002 | 0.031 | –0.028 | –0.025 | 0.065 |
| *Household_Income* | –0.031 | 0.073 | –0.016 | 1.000 | 0.017 | 0.046 | 0.123 | 0.156 | 0.117 |
| *Household_Education* | –0.032 | 0.070 | 0.002 | 0.248 | 1.000 | 0.037 | –0.042 | –0.050 | 0.095 |
| *Connection_Speed* | 0.058 | 0.010 | 0.031 | 0.046 | 0.054 | 1.000 | 0.010 | 0.038 | –0.058 |
| *Child_Present* | 0.046 | –0.044 | –0.028 | 0.123 | 0.017 | 0.010 | 1.000 | 0.640 | 0.042 |
| *Household_Size* | 0.044 | –0.043 | –0.025 | 0.156 | 0.088 | 0.038 | 0.640 | 1.000 | –0.001 |
| *Household_Oldest_Age* | 0.003 | 0.022 | 0.065 | 0.117 | –0.055 | –0.058 | 0.042 | –0.001 | 1.000 |

*(continues)*

Table 5. Continued

Panel B. Descriptive statistics for computer hardware category

| Variables | Mean | Number | Standard deviation | Maximum | Minimum |
|---|---|---|---|---|---|
| $N_{ij}$ | 3.26 | 3,322 | 1.82 | 20 | 1 |
| $\log(p_{ij})$ | 4.43 | 3,322 | 1.87 | 11.05 | 0 |
| $N_{i,j-1}$ | 0.38 | 3,322 | 1.60 | 17 | 0 |
| *Household_Income* | 4.46 | 3,322 | 1.66 | 7 | 1 |
| *Connection_Speed* | 0.48 | 3,322 | 0.50 | 1 | 0 |
| *Household_Education* | 2.85 | 2,420 | 1.41 | 5 | 0 |
| *Household_Oldest_Age* | 6.82 | 3,322 | 2.60 | 11 | 1 |
| *Child_Present* | 0.48 | 3,322 | 0.50 | 1 | 0 |
| *Household_Size* | 3.07 | 3,322 | 0.35 | 6 | 1 |

Correlation matrix for computer hardware category

| | $N_{ij}$ | $\log(p_{ij})$ | $N_{i,j-1}$ | *Household_ Education* | *Household_ Income* | *Connection_ Speed* | *Household_ Size* | *Household_ Oldest_Age* | *Child_ Present* |
|---|---|---|---|---|---|---|---|---|---|
| $N_{ij}$ | 1.000 | 0.085 | 0.237 | –0.013 | 0.053 | 0.072 | 0.027 | –0.059 | 0.040 |
| $\log(p_{ij})$ | 0.085 | 1.000 | –0.041 | 0.072 | 0.115 | 0.047 | 0.102 | –0.121 | 0.077 |
| $N_{i,j-1}$ | 0.237 | –0.041 | 1.000 | –0.012 | 0.033 | 0.009 | –0.020 | –0.047 | –0.003 |
| *Household_Education* | 0.006 | 0.017 | 0.018 | 1.000 | 0.270 | 0.041 | 0.003 | 0.058 | –0.001 |
| *Household_Income* | 0.053 | 0.115 | 0.033 | –0.018 | 1.000 | 0.053 | 0.132 | 0.074 | 0.100 |
| *Connection_Speed* | 0.072 | 0.047 | 0.009 | 0.004 | 0.053 | 1.000 | 0.027 | –0.112 | 0.004 |
| *Household_Size* | 0.027 | 0.102 | –0.020 | 0.072 | 0.132 | 0.027 | 1.000 | –0.079 | 0.665 |
| *Household_Oldest_Age* | –0.059 | –0.121 | –0.047 | –0.074 | 0.074 | –0.112 | –0.079 | 1.000 | –0.053 |
| *Child_Present* | 0.040 | 0.077 | –0.003 | 0.036 | 0.100 | 0.004 | 0.665 | –0.053 | 1.000 |

Panel C. Descriptive statistics for air travel category

| Variables | Mean | Number | Standard deviation | Maximum | Minimum |
|---|---|---|---|---|---|
| $N_{ij}$ | 3.31 | 7,654 | 2.27 | 19 | 1 |
| $\log(p_{ij})$ | 5.71 | 7,654 | 1.05 | 10.85 | 0 |
| $N_{i,j-1}$ | 0.82 | 7,654 | 1.92 | 15 | 0 |
| *Household_Income* | 4.76 | 7,654 | 1.66 | 7 | 1 |
| *Connection_Speed* | 0.48 | 7,654 | 0.50 | 1 | 0 |
| *Household_Education* | 3.12 | 5,457 | 1.37 | 5 | 0 |
| *Household_Oldest_Age* | 6.52 | 7,654 | 2.59 | 11 | 1 |
| *Child_Present* | 0.43 | 7,654 | 0.49 | 1 | 0 |
| *Household_Size* | 2.98 | 7,654 | 1.37 | 6 | 1 |

Correlation matrix for air travel category

| | $N_{ij}$ | $\log(p_{ij})$ | $N_{i,j-1}$ | *Household_ Education* | *Household_ Income* | *Connection_ Speed* | *Child_ Present* | *Household_ Size* | *Household_ Oldest_Age* |
|---|---|---|---|---|---|---|---|---|---|
| $N_{ij}$ | 1.000 | 0.151 | 0.141 | 0.000 | 0.011 | 0.045 | 0.025 | 0.010 | 0.023 |
| $\log(p_{ij})$ | 0.151 | 1.000 | 0.018 | 0.023 | 0.044 | 0.022 | –0.019 | –0.002 | 0.002 |
| $N_{i,j-1}$ | 0.141 | 0.018 | 1.000 | –0.005 | 0.055 | 0.046 | 0.004 | 0.012 | 0.012 |
| *Household_Education* | 0.042 | 0.031 | 0.056 | 1.000 | 0.247 | 0.009 | 0.010 | –0.013 | 0.077 |
| *Household_Income* | 0.011 | 0.044 | 0.055 | –0.018 | 1.000 | 0.050 | 0.096 | 0.126 | 0.188 |
| *Connection_Speed* | 0.045 | 0.022 | 0.046 | 0.011 | 0.050 | 1.000 | 0.007 | 0.052 | –0.075 |
| *Child_Present* | 0.025 | –0.019 | 0.004 | 0.040 | 0.096 | 0.007 | 1.000 | 0.653 | 0.047 |
| *Household_Size* | 0.010 | –0.002 | 0.012 | 0.078 | 0.126 | 0.052 | 0.653 | 1.000 | 0.000 |
| *Household_Oldest_Age* | 0.023 | 0.002 | 0.012 | –0.077 | 0.188 | –0.075 | 0.047 | 0.000 | 1.000 |

Table 6. Regression Results (Dependent Variable: $N_{ij}$)

| Independent variables | Music | | Computer hardware | | Air travel | |
|---|---|---|---|---|---|---|
| $\log(p_{ij})$ | 0.01 | 0.01 | 0.15*** | 0.11**** | 0.29**** | 0.32**** |
| | (0.01) | (0.01) | (0.06) | (0.02) | (0.03) | (0.02) |
| $N_{i,j-1}$ | 0.34**** | 0.31**** | 0.42**** | 0.37**** | 0.17**** | 0.16**** |
| | (0.02) | (0.01) | (0.03) | (0.03) | (0.02) | (0.01) |
| *Connection_Speed* | 0.11*** | 0.12*** | 0.16**** | 0.31**** | 0.14** | 0.18*** |
| | (0.04) | (0.04) | (0.09) | (0.08) | (0.06) | (0.05) |
| *Household_Income* | –0.03* | –0.03** | 0.05* | 0.05** | –0.04** | –0.02 |
| | (0.01) | (0.01) | (0.03) | (0.03) | (0.02) | (0.01) |
| *Household_Education* | –0.02 | | –0.02 | | 0.05*** | |
| | (0.02) | | (0.03) | | (0.02) | |
| *Household_Oldest_Age* | –0.01 | –0.01 | –0.03* | –0.03* | 0.02 | 0.02** |
| | (0.01) | (0.01) | (0.02) | (0.02) | (0.01) | (0.01) |
| *Child_Present* | 0.15** | 0.10** | –0.03 | 0.18* | 0.21*** | 0.17*** |
| | (0.06) | (0.05) | (0.14) | (0.11) | (0.08) | (0.07) |
| *Household_Size* | 0.01 | 0.02 | 0.07 | –0.02 | –0.01 | –0.03 |
| | (0.02) | (0.02) | (0.05) | (0.04) | (0.03) | (0.02) |
| Intercept | 1.91**** | 1.84**** | 3.34**** | 3.40**** | 1.28**** | 1.19**** |
| | (0.12) | (0.09) | (0.27) | (0.21) | (0.21) | (0.17) |
| *N* | 2,695 | 3,851 | 2,420 | 3,322 | 5,457 | 7,654 |
| *R*-square | 0.118 | 0.102 | 0.091 | 0.073 | 0.046 | 0.045 |
| Adj. *R*-square | 0.115 | 0.100 | 0.087 | 0.070 | 0.044 | 0.044 |

**** $p < 0.001$; *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

all the three categories with the $Household\_Education_i$ variable in the smaller sample and without the $Household\_Education_i$ variable in the larger sample separately. The results from these two sets of regressions for each category are presented in Table 6.

To address the multicollinearity problem, we checked the condition index (CI) and the tolerance of each variable. The condition indices of the principal components are all less than 25, and the tolerances of the variables all exceed 0.1. There is no evident problem of multicollinearity in our regression model.

The empirical results supported most of our hypotheses and also provide some other interesting results. We summarize the results of the empirical hypotheses testing in Table 7 and illustrate their implications below.

The connection speed and search propensity variables are positively correlated with search depth across all of the three categories. So, Hypotheses 1 and 3 are robustly supported. Connection speed is closely related to the cost of conducting search online. Those households that adopted broadband Internet connection have lower search costs; therefore, based on Proposition 1, they tend to search more actively. Those consumers who searched more in the last session are identified as active “searchers,” with attributes of a consumer with high search depth such as Internet experience, high preference to quality, patience, and so on. These attributes are inherent to each consumer. Therefore, they can predict their search depth in the next session. The increasing adoption of broadband Internet connection and the more skilled consumers in

Table 7. Hypotheses Testing Results

| | Variables in the empirical model | Hypothesized correlation with search depth | Music | Computer hardware | Air travel |
|---|---|---|---|---|---|
| H1 | Broadband Internet adoption | Positive | Supported | Supported | Supported |
| H2 | Price | Positive | Not significant | Supported | Supported |
| H3 | Lag of search depth | Positive | Supported | Supported | Supported |
| H4 | Household income | Positive | Not supported | Supported | Not supported |
| H5 | Household education | Positive | Not significant | Not significant | Supported |
| H6 | Household age | Positive | Not significant | Not supported | Partially supported |

online searching explain our observation of greater search depth compared with Johnson et al. [14].

Hypothesis 2 is partially supported. The price of purchased goods significantly positively affects the search depth for computer hardware and air travel products but has no significant effect for music CDs. For high-value products (i.e., computer hardware and airline tickets in our sample), price alone is an important measure when a consumer makes search or purchase decisions. Yet, when consumers shop for relatively lower-value products such as music CDs, price is not a significant factor affecting their search. They may put more consideration into the brand name and service quality of the online retailers, as suggested in Smith and Brynjolfsson [23].

High-income consumers have high search costs, relatively low price sensitivity, and high quality preference. As analyzed above, income has a double effect on consumers' search—cost effect and quality effect. The relative strength of the two opposite effects leads to mixed results about the association of income and search depth on the Internet across categories. For music and travel products, high-income households search less but the result is reversed for computer hardware products. It is probably due to the higher differentiation level with respect to the services for computer products (quality guarantee, return and warranty, delivery and reliability, etc.) than for music CDs. As a result, the quality preference effect outweighs the increased search cost for high-income households in the computer hardware category, but the effects are reversed for music CDs. For air travel tickets, even though the service quality differentiation is also very high, we observe an inverse relationship between income and search depth. We suspect this is because the price–quality correlation among airlines is higher than among computer vendors on the Internet, and high-income households care less about the price. According to Corollary 1, other things being equal, high price–quality correlation reduces search.

Hypothesis 5 is supported only for the air travel category: higher-educated households tend to search more air travel Web sites before purchasing.

Older households are found to search less for computer hardware products but search more for air travel tickets. So Hypothesis 6 is partially supported. Households with older members may value the conveniences and comfort of traveling more than the younger ones. So they tend to search more airline Web sites for a better fit. Given the high degree of differentiation in the air travel services, these households obtain more benefits for conducting an additional search than the younger ones. Therefore, these households demonstrate greater search depth in the air travel category. On the other hand, age is related with experiences and familiarity with the product category and information-processing ability, which may decrease their search cost and therefore reduce their tendency to search. Younger customers are far more Internet savvy and are likely to search more given their familiarity with the Web. This explains the finding of search depth in the computer hardware category.

Other interesting results from our work include that households with children are found to search more actively among all categories, and that household size has no significant effect on search depth. Households with children may have a lower tolerance to risk, and therefore they have a higher demand for service and quality. According to Proposition 1, these households will search more. The three types of goods can be easily shared among the household members, and their searching and purchasing activities are substitutable. Therefore, with our data sample, we do not observe household size significantly associated with household search depth.

## Discussion and Conclusion

### Summary of Findings

THIS RESEARCH THEORETICALLY ANALYZES consumers' search behavior on the Internet and empirically tests the factors that affect search depth. It has great implications for business researchers and practitioners to understand consumers' online search activities and to choose competitive strategies accordingly.

Even though the general merchandiser Web sites such as Yahoo stores, eBay.com, and Walmart.com were eliminated from the list of online stores to avoid double counting, which may undermine our estimates of search depth, we see a significant increase in consumers' search over the Web from what was reported in Johnson et al. [14]. The two results were derived from consumer online shopping data collected from the same source and in similar categories, but before and after the boom-and-bust stages of e-commerce. Removing the noises in those "irrational" stages, we can see the underlying growth of e-commerce. This interesting and meaningful comparison sheds light on changes of the firms, markets, and consumers with the development of e-commerce: more online businesses appear in the online marketplace; the online market becomes competitive and consumers are less loyal to a particular Web retailer than at the early stage. On the other hand, consumers are found to take service quality of the Web stores into consideration besides prices when making purchase decisions, which mitigates price competition and encourages online retailers to differentiate themselves through services.

By incorporating both price and nonprice attributes into the classic search model, we build an analytical model to solve consumers' optimal online search strategy. The model suggests that search cost is inversely correlated with search depth while consumers' quality preference is positively correlated with search depth. Moreover, the distribution of price and quality offered by stores also influences search depth. We find that with some restrictions on the distribution functions, higher mean or higher variance of quality-adjusted price increases search depth. Verifying the above predictions with the household clickstream data, we find significant connections between consumers' search depth and cost of search, consumer characteristics (connection speed, search propensity, income, education, age, and presence of children), and product characteristics (mean and variance of price and quality distribution), which can be explained by theories from our analytical model.

The empirical findings support and extend the conclusions of the previous works on offline search. Searching the Internet requires technology support and knowledge of information technology. In this sense, connection speed is a unique factor in online search; education and age are provided with new interpretations and they are found with different effects on search depth from that in the traditional market.

This study also raises some closely related questions, such as how the cognitive cost affects consumer search, how price comparison Web sites attract different types of consumers and affect their search behaviors, and so on. Those are interesting issues that need to be addressed in future studies.

## Theoretical Contributions and Implications

Our analytical model extends the traditional one-dimensional search model by incorporating both price and nonprice attributes. It has several important implications. First, recent empirical evidence ([23] and empirical results in this paper) shows that consumers do give significant weight on quality factors such as brands, reputation, service quality, and privacy policy in making purchase decisions. For example, Smith and Brynjolfsson [23] showed that on average consumers are willing to pay $2.45 more for a particular item at Amazon because of its more trusted brand name. Therefore, adding the quality variable into consumers' search model is closer to consumers' real search activities. It allows us to analyze the roles of quality attributes on consumers' search and better understand consumers' search and purchase decision.

Second, the Internet has significantly reduced the cost to consumers to search for product information, especially for pricing information. When price becomes transparent, consumers were expected to purchase from the seller with the lowest price by Bertrand competition. Therefore, reduced search costs have been considered a threat to the online retailers' profit margin. Now that quality is also found to be a significant factor that consumers use to evaluate stores, retailers are relieved from the fierce price competition. They can provide varying prices for different levels of service quality or brand names.

Finally, adding quality in the analysis allows us to derive new insights on consumer search. In particular we can draw conclusions about how quality uncertainty and correlation of price and quality affect search depth from our analytical model.

## Practical Implications

Research results presented in this paper have the following practical implications for designing online retail Web sites and their underlying information systems. First, we find that consumers tend to search more and to be less loyal to a particular online retailer over time. This finding posits both the challenge of retaining existing customers and the opportunity of attracting new customers for online retailers. Hence, it is crucial to design online retail systems that can effectively anticipate and fulfill every customer's needs. To meet this requirement, recent technology advancement in personalization technologies [19] and recommender systems [1] should be incorporated into the design of online retail systems.

Second, we find that consumers consider product price as well as service quality when making purchase decisions. This finding suggests that online retailers design online retail systems flexible enough that consumers can customize service offers (e.g., warranty terms and delivery options) associated with products purchased.

Finally, this paper has shown that using automatically collected Web logs is a promising way to study consumers' online behaviors. It suggests that online retailers analyze Web logs using statistical tools, data-mining methods, and machine-learning algorithms. The analysis results can be used to understand consumers' online navigation and purchasing behaviors and to predict their future behaviors. The analysis results could be embedded into online retail systems.

*Acknowledgments:* The authors thank the reviewers and the editors for their constructive comments and helpful suggestions on earlier versions of the paper. They also thank the Global Knowledge Management Center at the University of Utah for access to ComScore Media Matrix data that was used in this paper.

## References

1. Adomavicius, G., and Tuzhilin, A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering, 17,* 6 (2005), 734–749.

2. Alba, J.; Lynch, J.; Weitz, B.; Janiszewski, C.; Lutz, R.; Sawyer, A.; and Wood, S. Interactive home shopping: Consumer, retailer, and manufacturer incentives to participate in electronic marketplaces. *Journal of Marketing, 61,* 3 (1997), 38–53.

3. Bakos, Y. Reducing buyer search costs: Implications for electronic marketplaces. *Management Science, 43,* 12 (1997), 1676–1692.

4. Beatty, S.E., and Smith, S.M. External search effort: An investigation across several product categories. *Journal of Consumer Research, 14,* 1 (1987), 83–95.

5. Bhatnagar, A., and Ghose, S. An analysis of frequency and duration of search on the Internet. *Journal of Business, 77,* 2 (2004), 311–330.

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

7. Brynjolfsson, E., and Smith, M. Frictionless commerce? A comparison of Internet and conventional retailers. *Management Science, 46,* 4, (2000), 563–585.

8. Bucklin, R.E.; Lattin, J.M.; Ansari, A.; Bell, D.; Coupey, E.; Gupta, S.; Little, J.D.C.; Mela, C.; Montgomery, A.; and Steckel, J. Choice and the Internet: From clickstream to research stream. *Marketing Letters, 13,* 3 (2002), 245–258.

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

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

11. Cole, C.A., and Balasubramanian, S.K. Age differences in consumers' search for information: Public policy implications. *Journal of Consumer Research, 20,* 1 (June 1993), 157–169.

12. Cooley, R.; Mobasher, B.; and Srivastava, J. Data preparation for mining World Wide Web browsing patterns. *Knowledge and Information Systems, 1,* 1 (1999), 1–27.

13. Donthu, N.; and Garcia, A. The Internet shopper. *Journal of Advertising Research, 39,* 3 (1999), 52–58.

14. Johnson, E.J.; Moe, W.W.; Fader, P.S.; Bellman, S.; and Lohse, J. On the depth and dynamics of World Wide Web shopping behavior. *Management Science, 50,* 3 (2004), 299–308.

15. Katona, G., and Mueller, E. A study of purchase decisions. In L.H. Clark (ed.), *Consumer Behavior: The Dynamics of Consumer Reaction.* New York: New York University Press, 1955, pp. 30–87.

16. Kiel, G.C., and Layton, R.A. Dimensions of consumer information seeking. *Journal of Marketing Research, 18,* 2 (May 1981), 233–239.

17. Kohn, M., and Shavell, S. The theory of search. *Journal of Economic Theory, 9,* 2 (October 1974), 93–123.

18. Moorthy, S.; Ratchford, B.T.; and Talukdar, D. Consumer information search revisited: Theory and empirical analysis. *Journal of Consumer Research, 23,* 4 (1997), 263–277.

19. Murthi, B.P.S., and Sarkar, S. The role of the management sciences in research on personalization. *Management Science, 49,* 10 (2003), 1344–1362.

20. Newman, J.R. Consumer external search: Amount and determinants. In A. Woodside, J. Sheth, and P. Bennet (eds.), *Consumer and Industrial Buying Behavior.* New York: Elsevier, 1977, pp. 79–84.

21. Ratchford, B.T.; Lee, M.S.; and Talukdar, D. The impact of the Internet on information search for automobiles. *Journal of Marketing Research, 40,* 2 (2003), 193–209.

22. Reinganum, J.F. A simple model of equilibrium price dispersion. *Journal of Political Economy, 87,* 4 (1979), 851–858.

23. Smith, M.D.; and Brynjolfsson, E. Consumer decision-making at an Internet shopbot: Brand still matters. *Journal of Industrial Economics, 49,* 4 (2001), 541–557.

24. Smith, M.D.; Bailey, J.; and Brynjolfsson, E. Understanding digital markets: Review and assessment. In E. Brynjolfsson and B. Kahin (eds.), *Understanding the Digital Economy: Data Tools and Research.* Cambridge: MIT Press, 2000, pp. 99–136.

25. Spiliopoulou, M.; Mobasher, B.; Berendt, B.; and Nakagawa, M. A framework for the evaluation of session reconstruction heuristics in Web-usage analysis. *INFORMS Journal on Computing, 15,* 2 (2003), 171–190.

26. Udell, J.C. Prepurchase behavior of buyers of electrical appliances. *Journal of Marketing, 30,* 4 (January 1966), 50–52.

27. Weitzman, M. Optimal search for the best alternative. *Econometrica, 47,* 3 (1979), 641–654.

# Appendix. Proofs to Propositions

## Proof to Proposition 1

APPLYING ENVELOPE THEOREM to Equation (6), we have

$$\frac{\partial E(N)}{\partial \theta} = \frac{f_M(m_r)}{F_M(m_r)^2} \frac{\partial m_r}{\partial \theta} < 0 \tag{A1}$$

$$\frac{\partial E(N)}{\partial c} = \frac{f_M(m_r)}{F_M(m_r)^2} \frac{\partial m_r}{\partial c}. \tag{A2}$$

Given the specification of the utility function, it is easy to see Equation (A1) is positive. That is, high-quality preference consumers tend to search more stores. To test the sign of Equation (A2), we take the derivative of both sides of Equation (5′) in terms of $c$, which yields

$$\frac{\partial m_r}{\partial c} = \frac{1}{F_M(m_r)} > 0. \tag{A3}$$

Therefore, if we replace $\partial m_r/\partial c$ in Equation (A2) with (A3), we have $(\partial E(N)/\partial c)$. Q.E.D.

## Proof to Proposition 2

If $F_M^1(m)$ first-order stochastic dominates $F_M^2(m)$, that is, $F_M^1(m) \leq F_M^2(m)$, then we have $E_1(m_r) \geq E_2(m_r)$ by the property of first-order stochastic dominance. Let $m_r^1$ and $m_r^2$ represent the stopping quality-adjusted prices under the two distribution functions $F_M^1(m)$ and $F_M^2(m)$, respectively. Because both $m_r^1$ and $m_r^2$ satisfy Equation (5′) with their own distribution functions $F_M^1(m)$ and $F_M^2(m)$, we have

$$\int_l^{m_r^1} F_M^1(m)\,dm - \int_l^{m_r^2} F_M^2(m)\,dm = 0$$

$$\int_l^{m_r^2} \left(F_M^2(m) - F_M^1(m)\right) dm + \int_l^{m_r^2} F_M^1(m)\,dm - \int_l^{m_r^1} F_M^1(m)\,dm = 0$$

$$\int_l^{m_r^2} \left(F_M^2(m) - F_M^1(m)\right) dm = \int_{m_r^2}^{m_r^1} F_M^1(m)\,dm.$$

Because $F_M^1(m) \leq F_M^2(m)$, then $\int_{m_r^2}^{m_r^1} F_M^1(m)dm \geq 0$ and $m_r^2 \leq m_r^1$. Hence, $E_1(N) \geq E_2(N)$. Q.E.D.

## Proof to Proposition 3

For two distribution functions $F_M^1(m)$ and $F_M^2(m)$ with the same mean and different variances $\sigma_1{}^2 \le \sigma_2{}^2$, then by the analysis of Kohn and Shavell [17], the stopping values have $m_r{}^2 \ge m_r{}^1$. Therefore, consumers are expected to search more stores when there is a higher utility uncertainty, that is, $E_1(N) \le E_2(N)$. Q.E.D.