Technology Usage and Online Sales: An Empirical Study

Prabuddha De, Yu (Jeffrey) Hu
Krannert School of Management, Purdue University, West Lafayette, Indiana 47907
{pde@purdue.edu, yuhu@purdue.edu}

Mohammad S. Rahman
Haskayne School of Business, University of Calgary, Calgary, Alberta T2N 1N4, Canada,
mohammad.rahman@haskayne.ucalgary.ca

Despite the widespread adoption of search and recommendation technologies on the Internet, empirical research that examines the effect of these technologies is scarce. How do online consumers use these technologies? Does consumers’ technology usage have an effect on the sales to them or their purchasing patterns? This paper empirically measures consumers’ usage of website technologies by analyzing server log data. We match technology usage data to sales data, controlling for consumers’ historical purchasing behavior. Our unique data set allows us to reveal the relationship between technology usage and online sales. Our analyses show that consumers’ information technology usage has a significant effect on the sales to them, but this effect varies for different technologies and across different products. In particular, the use of directed search has a positive effect on the sales of promoted products, whereas it has a negative effect on the sales of nonpromoted products. In contrast, the use of a recommendation system has a positive effect on the sales of both promoted and nonpromoted products. Surprisingly, the use of nondirected search has an insignificant effect on online sales.

Key words: electronic commerce; Internet; technology usage; online sales; search; recommendation

History: Received March 20, 2009; accepted June 15, 2010, by Lorin Hitt, information systems. Published online in Articles in Advance October 11, 2010.

1. Introduction

It’s increasingly important for e-commerce sites to use recommendations, in addition to advanced search and navigation, to help improve stickiness and increase revenue.

—Justin Ziegler, Cofounder & Head of IT, PriceMinister (Business Wire 2007)

Today’s Internet consumers face a dizzying array of product choices. For instance, Amazon.com offers every book in print, which means more than three million titles in the book category alone (Brynjolfsson et al. 2003). Visitors to Zappos.com, a leading online retailer of shoes, can choose from more than three million products across 1,000 brands (Greco 2007). To help consumers find the products they want from this sea of choices, virtually all Internet retailers have started to provide advanced technological features such as search functions and recommendation systems on their websites. Moreover, most Internet retailers plan to continue to invest in improving their search and recommendation technologies, according to a survey by Forrester Research (Mulpuru 2008).

Search and recommendation technologies, if used properly, can significantly enhance consumers’ shopping experience by reducing the steps required to locate the products for which consumers already have information and by allowing consumers to discover relevant products that they may not have sought out otherwise. In other words, these technologies not only influence how consumers utilize their prior product knowledge but also alter how they search for product information that is entirely new to them.

Even though search and recommendation technologies have become critical components of retailing websites (Business Wire 2007) and anecdotal evidence suggests that these technologies can lead to significantly higher revenues and profits for Internet companies (Siwicki 2007), empirical research that quantifies the economic impact of these technologies is surprisingly scarce. Most research to date has only focused on how to improve the design of these technologies (e.g., Ansari et al. 2000, Bodapati 2008). Researchers and practitioners alike seem to know very little about how consumers use such technologies and whether consumers’ usage of these technologies can have an effect on the sales to them or their purchasing patterns. In this paper, we study these largely ignored questions related to Internet commerce.

There are at least three reasons for the lack of research on the relationship between technology
usage and online sales. First, it is partly because of the difficulty of measuring consumers’ technology usage. Measuring a consumer’s usage of search and recommendation technologies requires researchers to track each of the hundreds and sometimes thousands of clicks made by a consumer during a typical shopping session (or a single visit to the website). Collecting and analyzing the clickstream data resulting from many shopping sessions can consume a huge amount of time and computing resources. Second, previous studies on consumers’ usage of certain website features have not been able to match the data on consumers’ technology usage with the data on the sales to them (Johnson et al. 2004). Limited by the availability of data, these studies only consider the correlation between the usage of these features and the possibility of consumers abandoning their shopping carts (e.g., Montgomery et al. 2004). Third, the difficulty of studying this topic also lies in the existence of consumer heterogeneity as a possible confounding factor—that is, a correlation between consumers’ technology usage and the sales to them may merely reflect the fact that consumer heterogeneity in preference for the firm’s products may be correlated with both variables. Therefore, to carefully reveal the relationship between technology usage and sales, researchers must control for consumer heterogeneity. But the data that capture a long history of purchases made by consumers so that consumer heterogeneity can be controlled for are difficult to obtain. As we describe below, we overcome all three difficulties in this paper.

Our research draws heavily upon the marketing literature on consumer behavior. This literature shows that consumers pass through an “information search” stage before the stages of “alternative evaluation” and “purchase” (Kotler 2002). During information search, consumers first try to activate prior knowledge stored in their memory and then try to acquire new information from external sources, including mass media, advertising, salespeople, family, and friends (Engel et al. 1990). Search and recommendation technologies implemented by Internet retailers can facilitate consumers’ recall of prior knowledge as well as their external search for additional product information, which in turn can lead to changes in their online purchases. In this paper, we use a unique and rich data set to examine this relationship between technology usage and online sales.

First, we directly measure consumers’ information technology (IT) usage by analyzing a retailing firm’s server log. This firm provides a search box and a recommendation engine on its website. Some consumers who purchase from its website never take advantage of these technologies; they simply browse through available products. In contrast, other consumers use the search and recommendation technologies provided by the firm. When consumers perform searches with exact product names or product stockkeeping units (SKUs), which we call “directed search” along the lines of Moe (2003), they are taken directly to the product page that displays the product being searched for. Thus, directed search enables consumers to effectively recall and utilize their prior knowledge. When consumers conduct searches with nonspecific keywords, which we call “nondirected search,” they are presented with a list of products that are relevant to the nonspecific keyword being searched for.

Unlike in the case of directed search, they are not directly taken to a specific product page in this case. Consumers can also respond to the products recommended by the retailer’s recommendation engine, discovering relevant products that they may not have sought out otherwise. Thus, consumers’ use of nondirected search and the recommendation system can facilitate their external information search.

Next, we match IT usage measures to online sales data and use several econometric models to study the relationship between IT usage and online sales for different types of products. Our unique data set enables us to eliminate consumer heterogeneity as a confounding factor, ensuring the reliability of our empirical results. Interestingly, we find that consumers’ use of the recommendation system has a positive effect on the online sales to them and that this positive effect extends to all types of products. In contrast, consumers’ use of directed search has a positive effect on online sales, but this positive effect only applies to “promoted products.” Promoted products appear in the company’s print advertising; as a result, consumers are likely to have usable prior knowledge on these products. When the dependent variable is switched to online sales of “nonpromoted products,” or products that do not appear in print advertising, the effect of consumers’ use of directed search is actually negative. We also find results that are somewhat counterintuitive: consumers’ use of nondirected search does not have a significant impact on the online sales to them for either promoted or nonpromoted products.

In this paper, we examine the effect of IT usage on online sales. One may, however, argue that online sales may simultaneously have an effect on IT usage. To directly address this simultaneity issue, we estimate a simultaneous equation system and find that our main results remain qualitatively unchanged. One may also argue that consumers’ directed search usage may indicate their strong interest in making purchases. Thus, consumers with directed search usage may be self-selected to behave differently from those without such usage. To control for this self-selection
effect, we follow Rosenbaum and Rubin's (1983) approach of propensity score matching and create a sample of consumers without directed search usage that matches with consumers with directed search usage on dimensions such as socioeconomic variables, historical purchasing measures, and browsing patterns. An analysis of these two matched samples demonstrates that our results are robust to considering such a self-selection effect. In addition to testing for simultaneity and self-selection, we conduct various other robustness checks as well and find that our results are robust with respect to all of them.

Understanding the effect of IT usage on online sales has important managerial implications. Our results demonstrate that consumers' usage of search and recommendation technologies on a firm’s website has a significant impact on their online purchases from the firm. In addition, the size and sign of this effect may vary for different types of technologies and across different types of products. Therefore, firms should develop customized marketing and promotion strategies based on consumers’ IT usage data. Moreover, firms should consider the resulting changes to their online sales when they decide to facilitate or discourage consumers’ usage of different technologies.

The rest of this paper is organized as follows. We review the relevant literature in §2, develop our hypotheses on how consumers’ usage of various website technologies may affect online sales in §3, and discuss our research design in §4. We then present our empirical results in §5 and test the robustness of our results in §6. This paper concludes with a discussion of our findings and some broader implications in §7.

2. Literature Review

Our research draws upon the existing marketing literature on consumer behavior. This literature shows that consumers must search for information from various sources before they can evaluate alternatives and make purchases (Engel et al. 1990, Kotler 2002). When conducting information search, consumers first scan their memory for relevant prior stored knowledge and previous experience; they then conduct external searches for information (Engel et al. 1990). Many papers have examined consumers’ information search using survey data. They have identified several factors that influence the extensiveness of consumers’ external search, including perceived benefit and cost, market environment, prior knowledge, and individual differences; see Beatty and Smith (1987) for a review of these papers.

More recently, researchers have started to use lab experiments to study how consumers conduct information search and how consumers’ information search affects their purchases. Hauser et al. (1993) examine how consumers allocate their time across different sources when searching for information and find that information can alter their intention to purchase. Lynch and Ariely (2000) find that when online stores lower search costs for quality information, consumers’ choices become more skewed toward high-priced, high-quality products.

Our paper is also related to the literature on electronic commerce that studies online consumers’ purchasing behavior through the analysis of clickstream data. Lohse and Spiller (1998) point out that many website design features can influence an online store’s sales, calling for more research on this topic. However, such research is still scarce. In particular, the mechanism by which the usage of website technologies affects online sales is not well understood. Mandel and Johnson (2002) find that consumers’ preferences can be influenced by manipulating page designs and that consumers can dynamically adapt to the changes at a website. By analyzing clickstream data, Moe (2003) categorizes visits made by consumers to online stores into four types: knowledge building visits, hedonic browsing visits, directed buying visits, and search/deliberation visits. Moe and Fader (2004) develop a model to predict the visit-to-purchase conversion probability and find a positive correlation between repeat visits and purchases. Our paper differs from these papers by focusing on Internet consumers’ use of search and recommendation technologies and investigating the relationship between technology usage and purchases.

Our research is closely related to an interesting paper by Sismeiro and Bucklin (2004). Using clickstream data from an online seller of cars, the authors find that what consumers do and what they are exposed to while visiting the site can strongly predict whether they will make purchases. Their paper represents an important first step toward understanding the effect of basic website features such as browsing functions, different content pages, and comparison matrices. However, as the authors themselves acknowledge, the site they study is different from most Internet retailing sites: this site does not offer mainstream products, nor does it offer search functions or recommendation systems. Most importantly, they do not observe the actual (i.e., realized) purchases. In contrast, we observe consumers’ actual purchases. Moreover, we study how the relationship between technology usage and online sales varies across different types of products—specifically, promoted and nonpromoted products.

This paper represents an early effort to examine the relationship between technology usage and online sales. Building upon the literature on consumers’ information search behavior, we point out that different website technologies—directed search,
nondirected search, and recommendation systems—can play different roles in changing consumers’ internal and external information search behavior. This allows us to hypothesize that consumers’ technology usage can affect their purchases in different ways and different directions, depending on the specific technology and the specific type of products involved. This paper’s results enrich and deepen our current understanding of how IT usage affects online sales.

3. Hypotheses

Our hypotheses are built on the marketing literature on consumer behavior, particularly consumers’ information search behavior when making purchase decisions. According to this literature, the consumer decision process can be divided into five stages: need recognition, information search, alternative evaluation, purchase, and outcome (Engel et al. 1990, Kotler 2002). Consumers search for information so that they can make better, more satisfying purchases. During the “information search” stage, consumers first conduct internal search by recalling prior knowledge; if this search proves inadequate, they then decide to acquire additional information from external sources (Engel et al. 1990). Consumers conduct internal search first simply because recalling prior knowledge in their memory is much less costly than conducting external search (Punj and Staelin 1983). This literature has found that the amount of “usable prior knowledge” and the accessibility of information in memory have a negative effect on the amount of external search consumers perform (e.g., Punj and Staelin 1983, Biehal and Chakravarti 1986). It has also found that consumers can obtain prior knowledge from their previous purchase and consumption experience as well as from firms’ advertising (Russo and Johnson 1980, Nelson 1974). Another factor that influences consumers’ external search behavior is the expected benefits and costs of additional searches. Rational consumers constantly weigh expected benefits against costs and will stop searching whenever expected benefits are lower than costs (Stigler 1961, Rothschild 1974). Next, we discuss how consumers’ use of search and recommendation technologies affects their information search behavior—internal search as well as external search.

3.1. Directed Search Usage

The Internet consumers studied in this paper can obtain prior knowledge either from their past purchase and consumption experience or from the company’s print advertising. The consumers who possess prior knowledge can actively perform specific searches by using product SKUs or exact product names. This type of specific search, or “directed search,” takes consumers directly to the product page that displays the product being searched for, thereby helping them locate a product for which they already have prior knowledge in just one click. In contrast, the consumers who do not use the directed search feature need to make many browsing clicks before they can locate a product for which they already have prior knowledge. Thus, directed search usage facilitates consumers’ internal search, improving the ease with which consumers can recall and use prior knowledge (Moe 2003). Consumers’ information search behavior can directly influence their purchases. As consumers perform more directed search and rely more on prior knowledge, they are likely to purchase more promoted products—products that appear in print advertisements and, hence, are more likely to be a part of consumers’ prior knowledge.1

The consumers who effectively use prior knowledge will conduct less external search (Punj and Staelin 1983). For instance, 60% of the consumers do not perform any external search when deciding where to get auto repair service; these consumers simply rely on their prior knowledge (Biehal 1983). Therefore, directed search usage, by facilitating consumers’ internal search, lowers their incentive to perform external search. Nonpromoted products do not appear in print advertisements and as a result are unlikely to be among the products for which consumers have prior knowledge. Unless they discover nonpromoted products through external search, these products are unlikely to be purchased. Therefore, the consumers who have a higher directed search usage are likely to purchase fewer nonpromoted products because of the substitution of external search by internal search.

In sum, consumers are likely to purchase more promoted products and fewer nonpromoted products when they use more directed search. This observation is captured in the following hypotheses.

**Hypothesis 1A (H1A).** Directed search usage has a positive effect on the sales of promoted products.

**Hypothesis 1B (H1B).** Directed search usage has a negative effect on the sales of nonpromoted products.

3.2. Nondirected Search Usage

When internal search does not provide an adequate amount of information for consumers to make decisions, consumers can seek more information from the company’s website.2 However, some consumers never use any advanced technology features. They simply

---

1 Please note that, in this paper, we do not relate consumers’ technology usage (e.g., directed search usage) to their purchase of a specific product. This is because both technology usage and sales are measured at the consumer level.

2 For example, consider a consumer who has prior knowledge on two dresses. If the consumer finds either of the two dresses satisfactory, she may not conduct any external search. If, however, she...
browse through available products on a website, just like they do in brick-and-mortar stores. Other consumers may use the search function to perform non-specific searches by typing in a keyword that is not a product SKU or an exact product name. Such searches are called “nondirected search,” and they typically lead to a list of products. Consumers may subsequently choose to obtain further details about any of the products on the list. They may also choose not to obtain information about any of the products on the list.

Interactive features provided by online websites, including the nondirected search feature, can lower consumers’ search costs and enable the screening of alternatives so that only those products which are best suited to their tastes are considered (Alba et al. 1997). Rothschild (1974) shows that lower search costs lead to more searches, fewer market breakdowns, and more sales. Consistent with this literature, we hypothesize that because of the lower search costs resulting from using nondirected search, consumers may perform more external search and discover a larger amount of desirable products. As consumers gather additional information and consider more products, the likelihood of finding a product that provides positive utility improves and the sales to these consumers may increase (Roberts and Lattin 1997). Thus, the consumers who perform nondirected search may purchase more products than those who do not perform nondirected search. Because the products that consumers can discover through nondirected search include both promoted and nonpromoted products, nondirected search usage is likely to increase the sales of promoted products and nonpromoted products alike.

**Hypothesis 2A (H2A).** Nondirected search usage has a positive effect on the sales of promoted products.

**Hypothesis 2B (H2B).** Nondirected search usage has a positive effect on the sales of nonpromoted products.

### 3.3. Recommendation System Usage

Another advanced technology feature consumers can use when conducting external search is the website’s recommendation system. While a product page is viewed, the website will display five products that are related to the focal product. If a consumer clicks on one of the products being recommended, it enables the consumer to discover and explore products that she otherwise may not have been aware of and, hence, provides new information (Mobasher et al. 2001). The use of the recommendation system lowers the costs incurred by consumers when searching for additional information (Brynjolfsson et al. 2006). Consistent with Rothschild (1974), we hypothesize that as a result of using the recommendation system, consumers may conduct more external search, discover a larger number of desirable products, and eventually make more purchases. Because the recommendation system can suggest both promoted and nonpromoted products, as long as the suggested products are related to the focal product currently being viewed by consumers, we hypothesize that recommendation system usage is likely to increase the sales of promoted products and nonpromoted products alike.

**Hypothesis 3A (H3A).** Recommendation system usage has a positive effect on the sales of promoted products.

**Hypothesis 3B (H3B).** Recommendation system usage has a positive effect on the sales of nonpromoted products.

To summarize, we expect the effect of technology usage on online sales to vary for different technologies and across different products. The next section describes our research setting and data.

### 4. Research Design

#### 4.1. Website Description

The data for our study come from a large retailer of women’s clothing. One may go to a category page of the company’s website and simply browse the products that are available under that product category, which could be tops, pants, sweaters, dresses, etc. Each product page shows a picture of a model wearing the product, as well as the price, available sizes, and colors. Consumers can easily move from one category to another and from one product to a different product. The process of browsing through the available products on this company’s website resembles the process of browsing through the products displayed on shelves in a brick-and-mortar store.

The retailer’s site also provides some advanced features such as a search function and a recommendation system. A website visitor may search for a product of interest by using the search function. If the visitor searches for a specific product, with either its SKU or the exact product name, the website takes her directly to the product page of that specific product. In contrast, if she searches with a nonspecific keyword, the website presents a list of relevant products that match the search keyword. In addition, when a

---

3 The retailer has requested to remain anonymous.

4 Of the 4,108 search terms used by consumers, 1,266 specified product SKUs or product names; searches using these terms are considered directed search. The other search terms (2,842) did not specify product SKUs or product names; searches using these other terms are considered nondirected search.
visitor views a product page, the website always recommends five other products that the retailer feels the visitor “may also like.” These five products are predetermined by the company’s experts based on their similarity with the focal product and are recommended to all those who view the focal product. We note that this type of nonpersonalized recommendation system is widely adopted by many Internet retailers on their product pages, including Macys.com, Blockbuster.com, and Amazon.com.\(^5\) The static nature of this site’s recommendation system alleviates any concerns regarding discrepancies in the list of recommended products across visitors. If a consumer clicks on one of the recommended products, she will be taken to the page of the clicked product. We measure consumers’ usage of the recommendation system by counting the number of times they click on one of the recommended products. This approach is similar to the counting of clicks in online advertising.

Once on the product page, a visitor can add the product to the shopping cart. Subsequently, the visitor may purchase the products in the current shopping cart or continue to shop. The visitor may also choose to login and manage her account. For each page requested by a visitor and each time the visitor clicks on a dynamic object such as the shopping cart, the company’s server log records the URL of the page, time stamp, cookie identification, and session identification.

4.2. Data

Our data set contains information of all orders placed through the company’s Internet channel (and also its catalog channel) from May 2003 to April 2006.\(^6\) For each item purchased from the company, we have information regarding the price paid, date of transaction, consumer’s unique identification, whether or not the item was returned, ordering channel (i.e., Internet or catalog), and purchase identification. Overall, we have the data for seven million purchases that were made by approximately one million unique consumers. The company promotes its products by regularly sending catalogs and e-mails to consumers. Each catalog contains a subset of the products being sold by this company. For each product, the catalog provides a picture of the model wearing the product, as well as the price, available colors and sizes, and product SKU. Each e-mail contains a screenshot of the company’s website and a link to the website; it does not contain any information on specific products. We have the information regarding which catalogs and e-mails each consumer received between January 2005 and April 2006. For each catalog that was sent to consumers, we know which products were promoted in that catalog. This allows us to distinguish promoted products from nonpromoted products, products that were not promoted in catalogs but still had positive sales. Finally, we also have server logs recorded at the company’s website from March 2006 to April 2006, capturing each page request made by each visitor during this time. These server logs follow the standard World Wide Web Consortium extended log file format. We have approximately 52 million lines of logs for these two months, amounting to about 850,000 page requests per day.

Our rich and unique data set allows us to determine the online sales to each consumer. We can also measure each consumer’s usage of website technologies. Next, we discuss how we select a sample of consumers for our study and how we match each consumer’s technology usage with the sales to her.

4.3. Sample

Matching each consumer’s technology usage with the sales to her requires the identification of the website sessions carried out by each consumer. Fortunately, when a consumer makes a purchase online, the same Web order identification is recorded in both the server log and the purchase database, enabling us to identify all the purchase sessions carried out by the consumers who made a purchase. This paper focuses on the consumers with at least one Internet purchase because we can match their technology usage with the sales to them. This approach of selecting consumers with at least one Internet purchase is consistent with the sample selection approach in the existing literature (e.g., Wu and Rangaswamy 2003).

We also identify the nonpurchase sessions carried out by consumers. This is because consumers’ nonpurchase sessions, which may be categorized as hedonic browsing or knowledge building sessions, can have implications for the subsequent purchase sessions (Moe 2003). Because the company’s server log records a unique cookie number for each consumer, we can use the cookie number in each consumer’s purchase sessions to locate all the nonpurchase sessions made by the same consumer.\(^7\) In this paper, we consider consumers’ Internet purchase data in April 2006. Along with all the purchase sessions in April 2006, we also include the sessions 30 days prior to the

\(^5\) Amazon.com also provides a “personalized recommendations” page, displaying the recommendations generated for a particular user.

\(^6\) The company has one physical store that accounts for a negligible percentage of overall sales. We do not have data on store sales. Our results are, however, robust to including or excluding the consumers who may have access to the store.

\(^7\) For those consumers who regularly clean their browser cookies, we can identify all the purchase sessions, but we may miss some of their nonpurchase sessions. Only 10% of Internet users, however, clean cookies on a daily basis (Peterson 2005).
purchase sessions, a grace period consistent with previous studies (e.g., Sismeiro and Bucklin 2004). This implies that the sessions we consider can go as far back as March 1, 2006.

As mentioned before, the company regularly sends catalogs and e-mails to consumers to promote its products. All consumers do not receive all catalogs, although all consumers receive all e-mails. To control for the effect of the company’s marketing actions, we select only those consumers who received all the catalogs. Our analysis of the data shows that the impact of a catalog typically lasts for about 30 days, which is consistent with the retailer’s past experience. Therefore, in order to be conservative, we select the consumers who received all the catalogs that were sent out by the company between February 1, 2006 (59 days prior to April 1, 2006) and April 30, 2006.

To summarize, we only consider those consumers who received every catalog sent out between February 1, 2006, and April 30, 2006, and made at least one Internet purchase in April 2006. We study the relationship between these consumers’ technology usage and the sales to them in April 2006. Their purchase data prior to April 1, 2006, is used to calculate historical purchasing measures, which act as control variables for consumer heterogeneity. Having a physical store in Florida forces the retailer to collect sales taxes on the sales to Florida consumers. We have excluded Florida consumers from our sample, thereby eliminating variations in sales taxes across consumers as a confounding factor, although our results remain the same when we include these consumers. Our final sample contains 8,199 consumers.

4.4. Main Variables
Our research design allows us to study how the effect of technology usage varies across different types of products, i.e., promoted products and nonpromoted products. We categorize a product as “promoted” if it was included at least once in the catalogs sent between February 1, 2006, and April 30, 2006; the rest of the products are categorized as “nonpromoted” products. The sales to each consumer are calculated as the dollar amount spent by the consumer, with the products that were returned netted out. The definition of all the variables can be found in Table 1. The first three rows in Table 2 present the descriptive statistics of the sales to consumers in our sample. During April 2006, the average dollar amount spent by a consumer in the sample was $90.40, of which $76.90 was accounted for by promoted products and $13.50 by nonpromoted products.

To measure consumers’ usage of search and recommendation technologies, we first note each consumer’s total number of page requests. We then count the number of times a consumer performed directed search, the number of times she performed non-directed search, and the number of times she clicked on one of the recommended products, and we normalize these three measures by the total number of page requests made by the consumer. After normalization, each of the three technology usage variables—Directed Search Usage, Nondirected Search Usage, and Recommendation System Usage—measures the percentage of page requests related to using a particular technology.8 Such a normalization procedure eliminates concerns stemming from the disparity in the total number of page requests across consumers.

The last three rows in Table 2 provide the descriptive statistics of consumers’ technology usage in our sample. We note that, during April 2006, a consumer in our sample carried out 2.3 sessions and made 144.3 page requests on average. Among the 8,199 consumers in the sample, 6,109 did not use directed search; the mean of Directed Search Usage for the remaining 2,090 consumers is 3.5%, and the mean of Directed Search Usage for all consumers together is 0.9%. Similarly, 6,793 consumers did not use non-directed search; the mean of Nondirected Search Usage for the remaining 1,406 consumers is 1.8%, and the mean of Nondirected Search Usage for all consumers together is 0.3%. In addition, 4,377 consumers did not use the recommendation system; the mean of Recommendation System Usage for the remaining 3,822 consumers is 4.2%, and the mean of Recommendation System Usage for all consumers together is 2.0%. There were 2,814 consumers who never used any of the technologies.

8 The correlation matrix of all the variables can be found in the appendix.

Table 1 Variable Definitions

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Variable Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sales of Promoted Products</td>
<td>Total dollar amount spent by a consumer in April 2006 on the products that were included at least once in the catalogs between February 1, 2006, and April 30, 2006</td>
</tr>
<tr>
<td>Sales of Nonpromoted Products</td>
<td>Total dollar amount spent by a consumer in April 2006 on the products that were not included in the catalogs between February 1, 2006, and April 30, 2006</td>
</tr>
<tr>
<td>Overall Sales</td>
<td>Total dollar amount spent by a consumer</td>
</tr>
<tr>
<td>Directed Search Usage</td>
<td>Number of times a consumer performed directed search: searches with exact product names or product SKUs, divided by the total number of page requests</td>
</tr>
<tr>
<td>Nondirected Search Usage</td>
<td>Number of times a consumer performed nondirected search: searches with nonspecific keywords, divided by the total number of page requests</td>
</tr>
<tr>
<td>Recommendation System Usage</td>
<td>Number of times a consumer clicked on one of the recommended products, divided by the total number of page requests</td>
</tr>
</tbody>
</table>
Historical purchasing measures are widely used as controls for consumer heterogeneity in both the marketing literature and industry practices (Anderson and Simester 2004). We use two historical purchasing measures as controls for consumer heterogeneity: a variable named Days Since Last Purchase, which is defined as the number of days from the date of the last purchase on the Internet to April 1, 2006 (in its natural log), and a variable named Historical Total Purchases, which is defined as the total dollar amount spent on the Internet prior to April 1, 2006 (in its natural log).

5. Empirical Analyses

5.1. Basic Model
We first use a regression model to examine the effect of consumers’ technology usage on the sales to them. One may argue that the variation in consumers’ technology usage simply reflects consumer heterogeneity—consumers who used these technologies at least once might also be more loyal to the company’s products and this may explain why they purchased more. Such a result could still be an interesting finding in itself, similar to the observation of Hitt and Frei (2002) that customers who use a bank’s Internet channel to bank tend to be better customers. However, in this paper, we try to isolate the effect of technology usage from that of consumer heterogeneity by adding controls for consumer heterogeneity to our analyses (and subsequently also by using a fixed effects model; see §6.1).

The model we estimate includes three technology usage variables and two control variables:

\[ Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 T_{3i} + \beta_4 T_{4i} + \beta_5 X_{3i} + \epsilon_i, \]

(1)

where \( Y_i \) denotes the sales of promoted products (or nonpromoted products) to consumer \( i \); \( X_{1i}, X_{2i}, \) and \( X_{3i} \) are, respectively, the directed search usage, nondirected search usage, and recommendation system usage by her; \( T_{3i} \) is the number of days since her last purchase; \( T_{4i} \) is the total of her historical purchases; and \( \epsilon_i \) is the random error. We note that the model in Equation (1) allows consumers to have positive a value on more than one technology usage variable, with the total effect as simply the sum of the effects of all the technology usage variables.

Columns (1) and (2) of Table 3 present the estimates of the regression model when the dependent variable is the sales of promoted products and the sales of nonpromoted products, respectively.\(^9\) \(10\) These results show that Recommendation System Usage has a significantly positive effect on the sales of both promoted products and nonpromoted products, supporting for H3A and H3B. The results also show that Directed Search Usage has a significantly positive effect on the sales of promoted products while having a significantly negative effect on the sales of nonpromoted products. Thus, we find support for H1A and H1B. Finally, the results reveal some surprising outcomes:

---

*\(^9\) The \( R^2 \) values in Table 3 may seem low. However, because we are measuring consumer level behavior, this is not surprising, nor does it affect our interpretation of the coefficients (Wooldridge 2002).*

*\(^10\) To examine the effect of any potential correlation between the two dependent variables, we have also estimated a seemingly unrelated regression (SUR) model, and its results are identical to those in Table 3.*
the coefficient on Nondirected Search Usage is statistically insignificant, whether we use the sales of promoted products or the sales of nonpromoted products as the dependent variable. Thus, H2A and H2B are not supported. Table 4 summarizes these empirical results.

We note that there was a small segment of consumers (614) who used both directed and nondirected search. We have included an interaction term between Directed Search Usage and Nondirected Search Usage and reestimated the model in Equation (1). Our results remain qualitatively the same even when this interaction term is included. We have also explored path dependency by examining how consumers transition from direct search to nondirected search and vice versa. In doing so, we have created two dummy variables: the first one for the consumers who used both types of search and the second one for those who only used both types of search but used directed search first. Our results remain qualitatively the same when these dummy variables are included. Furthermore, we have considered—along the lines of Ghose and Yang (2009)—whether nondirected search with a large number of keywords differs from such a search with a small number of keywords. We have created a dummy variable that indicates whether the average length of nondirected search keywords used by a consumer is above the population median and have used it to separate Nondirected Search Usage into two variables—Nondirected Search Usage Short and Nondirected Search Usage Long. We find that replacing Nondirected Search Usage with these two new variables does not change our results. A small segment of consumers (482) also purchased from the catalog channel in addition to purchasing from the Internet channel. We have used consumers’ purchases from the catalog channel as a control variable, and our results are robust to adding this control variable. Finally, our results remain qualitatively the same when we control for consumers’ browsing patterns (number and duration of sessions, number of category pages viewed).

5.2. Considering Potential Simultaneity
One may argue that the effect of technology usage on sales we have found stems really from the effect of sales on technology usage. Even though sales “happen” after technology usage, they may still have an effect on technology usage in the following scenario: the consumers who have strong intentions to purchase may have higher incentives to use technology features, and these consumers may indeed purchase more in the end. Such a scenario leads to the simultaneity between sales and technology usage—they may influence each other at the same time. A classical example of simultaneity is the relationship between price and quantity: quantity responds to price changes (demand curve), whereas price is set according to quantity (supply curve). The appropriate econometric technique for such a scenario is to estimate a system of simultaneous equations (Greene 2002). Accordingly, we add to our original Equation (1) three more equations that describe the effects of sales on the three types of technology usage:

\[ X_{1i} = \gamma_{10} + \gamma_{11} Y_i + \eta_{1i}, \]  
\[ X_{2i} = \gamma_{20} + \gamma_{21} Y_i + \eta_{2i}, \]  
\[ X_{3i} = \gamma_{30} + \gamma_{31} Y_i + \eta_{3i}, \]

where \( \eta_{1i}, \eta_{2i}, \) and \( \eta_{3i} \) are the random errors in the three technology usage variables.

We adopt an approach developed by Hausman (1983) and Hausman et al. (1987) to estimate this system of simultaneous equations.\(^{11}\) This approach relies on instrumental variables as well as covariance restrictions. We use one instrumental variable \( Z_i \), which denotes the sales to consumer \( i \) prior to April 2006, and assume that \( \eta_{1i}, \eta_{2i}, \) and \( \eta_{3i} \) are uncorrelated with \( e_i \).\(^{12}\) First, we use an instrumental variable estimator to obtain consistent and unbiased estimates of the coefficients in Equations (2)–(4). The variable \( Z_i \) is logically uncorrelated with consumers’ technology usage in April 2006 \( (X_{1i}, X_{2i}, X_{3i}) \) and correlated with the sales to them in April 2006 \( (Y_i) \). We create the estimated residuals in Equations (2)–(4) and call them \( \hat{\eta}_{1i}, \hat{\eta}_{2i}, \) and \( \hat{\eta}_{3i} \). Next, we use \( \hat{\eta}_{1i}, \hat{\eta}_{2i}, \) and \( \hat{\eta}_{3i} \) as instrumental variables and apply an instrumental variable estimator to estimate Equation (1).

\(^{11}\) We thank Jerry Hausman for suggesting to us the estimation approach used here to consider potential simultaneity. Please refer to Wooldridge (2002, p. 227) and Greene (2002, p. 395) for details on this approach.

\(^{12}\) An example of the random factors captured by \( \eta_{1i}, \eta_{2i}, \) and \( \eta_{3i} \) is the possibility of recommendations and search results being slow to display on consumers’ computers at certain times or under certain conditions. Our assumption means that these random factors do not directly influence sales. They may still indirectly influence sales via the technology variables as shown in Equation (1), but they are orthogonal to the random error \( e_i \) in Equation (1).
Table 5 The Effect of Technology Usage, Using IV Estimator to Consider Potential Simultaneity

<table>
<thead>
<tr>
<th>Variable</th>
<th>(1) Sales of promoted products</th>
<th>(2) Sales of nonpromoted products</th>
</tr>
</thead>
<tbody>
<tr>
<td>Directed Search Usage</td>
<td>449.012**</td>
<td>-149.190**</td>
</tr>
<tr>
<td></td>
<td>(53.511)</td>
<td>(12.281)</td>
</tr>
<tr>
<td>Nondirected Search Usage</td>
<td>165.852</td>
<td>-44.684</td>
</tr>
<tr>
<td></td>
<td>(94.832)</td>
<td>(27.553)</td>
</tr>
<tr>
<td>Recommendation System Usage</td>
<td>189.013**</td>
<td>71.974**</td>
</tr>
<tr>
<td></td>
<td>(28.074)</td>
<td>(11.325)</td>
</tr>
<tr>
<td>Days Since Last Purchase</td>
<td>5.840**</td>
<td>0.193</td>
</tr>
<tr>
<td></td>
<td>(0.758)</td>
<td>(0.292)</td>
</tr>
<tr>
<td>Historical Total Purchases</td>
<td>5.532**</td>
<td>1.218**</td>
</tr>
<tr>
<td></td>
<td>(0.538)</td>
<td>(0.172)</td>
</tr>
<tr>
<td>Intercept</td>
<td>14.589</td>
<td>7.695**</td>
</tr>
<tr>
<td></td>
<td>(5.674)</td>
<td>(2.083)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.035</td>
<td>0.027</td>
</tr>
<tr>
<td>Sample size</td>
<td>8,199</td>
<td>8,199</td>
</tr>
</tbody>
</table>

Note. Robust standard errors are in parentheses. *p < 0.05; **p < 0.01.

5.3. Interpreting the Results

It is important to assess the economic significance of the results in Table 3. As mentioned in §4.4, the mean of Directed Search Usage is 3.5% across those consumers who used directed search at least once, whereas the mean of Directed Search Usage is zero for the other consumers. Using the coefficients on Directed Search Usage in Table 3 (490.536 and -143.413), we find that, everything else being equal, an average consumer who used directed search at least once purchased $17.20 (or 22.0%) more of promoted products and $5.00 (or 37.0%) less of nonpromoted products, compared with an average consumer who never used directed search. Similarly, from §4.4, the mean of Recommendation System Usage is 4.2% across those consumers who used the recommendation system at least once. Thus, using the coefficients on Recommendation System Usage in Table 3 (184.428 and 71.036), we estimate that, everything else being equal, an average consumer who used the recommendation system at least once purchased $7.75 (or 10.0%) more of promoted products and $2.98 (or 22.2%) more of nonpromoted products, compared with an average consumer who never used the recommendation system. Given that about 46.6% of the consumers used the recommendation system at least once and that a consumer spent $90.40 on average in April 2006, we conclude that the presence of the recommendation system increased the monthly sales of this retailer by more than 5.5%.

Our empirical results are consistent with our hypotheses, when we study the effect of online consumers’ usage of directed search and the recommendation system. First, we find that consumers’ usage of directed search has a positive effect on their purchase of promoted products and a negative effect on their purchase of nonpromoted products. Our explanation is that the directed search technology helps consumers conduct internal search by recalling prior knowledge with ease. Because consumers are likely to have prior knowledge on promoted products, the sales of promoted products increase. However, as consumers rely more on prior knowledge and internal search, they have less incentive to conduct external search. Thus, the sales of nonpromoted products decline.

Second, we find that consumers’ usage of the recommendation system has a positive effect on the sales of both promoted and nonpromoted products. Our finding at the consumer level is consistent with the finding of recent research (e.g., Chen et al. 2004) at the aggregate website level. Our explanation is that the company’s recommendation system can facilitate consumers’ external search, leading to more sales for both promoted and nonpromoted products.

However, when we study consumers’ usage of nondirected search, the results are surprising and inconsistent with our hypotheses. Conceptually, the usage of nondirected search could have helped consumers conduct more external search. This could have translated to higher sales. Instead, we find consumers’ usage of nondirected search is not correlated with their purchase of either promoted products or nonpromoted products. A plausible explanation is that

---

13 We thank the department editor for his suggestion to show the value of the recommendation system in terms of the gain in sales.
14 One could potentially use the total number of page requests made by each consumer as a proxy for the amount of newly acquired information. The mean number of page requests for the consumers who used directed search is 134.44, significantly smaller than that for those who did not use directed search (147.72). A t-test shows that this difference has a t-statistic of 3.16 and a p-value of less than 0.01.
15 The mean number of page requests for the consumers who used nondirected search (201.60) is significantly higher than that for those who did not (88.13), with a t-statistic of 35.23 and a p-value less than 0.01.
nondirected search might not have led to information that could be easily processed and used by consumers. Nondirected search typically produces a long list of products, providing a large amount of information, but the list of products is not ranked using any criterion that reveals the most relevant pieces of information to consumers. However, consumers have limited abilities to process information, as described by “the magical number seven, plus or minus two” (Miller 1956). They may incur high cognitive costs when presented with a large amount of information that is not easily processable (Simon 1955). Because of the information overload problem and the lack of a context for the information being provided, it is difficult for consumers to process all the information provided by nondirected search. The following quote from Simon (1971, pp. 40–41) may explain this better: “What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention, and a need to allocate that attention efficiently among the overabundance of information sources that might consume it.” We note that our explanation above is consistent with the marketing literature on information overload that has found that large amounts of overloaded information can adversely affect consumers’ decision making; see Jacoby (1984) for a review of this literature.

In contrast, consumers can effectively process and utilize the additional information provided by the recommendation system. This is because the recommendation system suggests a small number of products (namely, five, which is below the magical number seven suggested by Miller 1956). In addition, the recommended products are related to the focal product the consumer is currently considering, providing a context and an environment in which consumers can utilize the external information being provided. Klein and Yadav (1989) and Simonson and Tversky (1992) show that the context and environment can influence how consumers utilize information and make choices. Thus, the information on a small set of recommended products that are related to the focal product can be processed and utilized efficiently by consumers.

5.4. Effect on the Proportion of Nonpromoted Products in Overall Sales

We have already seen that recommendation system usage improves the sales of both promoted and nonpromoted products. From the results in Table 3, we cannot, however, easily infer how recommendation system usage influences consumers’ purchasing patterns (i.e., their relative tendency to purchase promoted products versus nonpromoted products). Toward this end, we calculate the fraction of nonpromoted products in overall sales for all consumers with positive sales (7,252 consumers). This variable is independent of overall sales, and we use a generalized linear model (GLM) to model it (Papke and Wooldridge 1996):

$$E(p_i \mid X_i) = \frac{e^{X_i\beta}}{1 + e^{X_i\beta}}, \quad 0 \leq p_i \leq 1,$$

(5)

where $p_i$ is the fraction of nonpromoted products in the overall sales to consumer $i$ in April 2006, and $X_i$ is a vector of explanatory variables that include $X_{i1}, X_{i2}, X_{i3}, T_{i1}, T_{i2}$.

We estimate the generalized linear model shown in Equation (5) and report the results in Table 6. The estimates show that the usage of directed search has a negative effect on the proportion of nonpromoted products in overall sales, whereas the usage of the recommendation system has a positive effect on the same proportion. Once again, the usage of nondirected search does not have a significant effect on the proportion of nonpromoted products in overall sales.

As mentioned before, directed search usage facilitates consumers’ internal search, lowering their incentive to acquire additional information from external sources. As a result, directed search usage skews consumers’ purchases toward the products for which consumers are likely to have prior knowledge, namely, promoted products. In contrast, recommendation system usage lowers the costs incurred by

---

17 A typical nondirected search on this company’s website would return dozens, sometimes hundreds, of products.

18 We regress the size of consumers’ consideration sets onto all the independent variables and find that the coefficient of Nondirected Search Usage is insignificant.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Percentage of nonpromoted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Directed Search Usage</td>
<td>-27.635$^{**}$</td>
</tr>
<tr>
<td>Nondirected Search Usage</td>
<td>-0.084</td>
</tr>
<tr>
<td>Recommendation System Usage</td>
<td>2.443$^{**}$</td>
</tr>
<tr>
<td>Days Since Last Purchase</td>
<td>-0.072$^{**}$</td>
</tr>
<tr>
<td>Historical Total Purchases</td>
<td>0.016</td>
</tr>
<tr>
<td>Intercept</td>
<td>-1.029$^{**}$</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>-3,053.9</td>
</tr>
<tr>
<td>Sample size</td>
<td>7,252</td>
</tr>
</tbody>
</table>

Note. Robust standard errors are in parentheses.

$p < 0.05$; $^{**}p < 0.01$. 

Table 6 The Effect of Technology Usage on the Proportion of Nonpromoted Products in Overall Sales
consumers when conducting external search, making them less reliant on prior knowledge (and internal search) and more reliant on external search. Note that a complete reliance on prior knowledge translates to almost zero purchases of nonpromoted products and that a complete reliance on external search leads to the purchase of both nonpromoted products and promoted products. Therefore, we expect recommendation system usage to increase the proportion of nonpromoted products in overall sales.\(^\text{19}\)

### 6. Robustness Checks

Section 5.2 shows that our findings survive when we consider potential simultaneity. Next, we demonstrate that they survive a wide range of other robustness checks as well.

#### 6.1. Using Fixed Effects Models to Control for Consumer Heterogeneity

So far, we have been using consumers’ historical purchasing measures to control for consumer heterogeneity. Now we check the robustness of our results by estimating a panel data model and using fixed effects to control for consumer heterogeneity. We organize our data on a subsample of the consumers who made more than one Internet purchase in March 2006 and April 2006 into a panel data set. For each consumer in this panel data set, we observe the sales and technology usage in each of her multiple purchase occasions.

We estimate the following fixed effects model:

$$Y_{ij} = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \beta_3 X_{3ij} + \beta_4 D_i + \epsilon_{ij}, \quad (6)$$

where \(Y_{ij}\) denotes the sales of promoted products (or nonpromoted products) to consumer \(i\) in purchase occasion \(j\); \(X_{1ij}, X_{2ij}, X_{3ij}\) are the technology usage variables for consumer \(i\) in purchase occasion \(j\); \(D_i\) is a set of dummy variables indicating individual fixed effects; and \(\epsilon_{ij}\) is the random error.

In our sample, we have 1,192 consumers who can be retained for the panel data analysis. We have an unbalanced panel data set with an average of 2.2 purchases by each consumer (panel). Columns (1) and (2) of Table 7 present the estimates of the fixed effects model when the dependent variable is the sales of promoted products and the sales of nonpromoted products, respectively. Reassuringly, these results are qualitatively similar to the results we have obtained in §5.1. In particular, both Directed Search Usage and Recommendation System Usage have a significantly positive effect on the sales of promoted products. Turning to the sales of nonpromoted products, we find that the effect of Directed Search Usage is negative and significant, consistent with the results in §5.1. The effect of Recommendation System Usage on the sales of nonpromoted products is still positive but is now insignificant at \(p < 0.05\). This is not surprising because fixed effects models tend to overestimate the standard errors, especially for a small number of panels (Allison 2005). Furthermore, the coefficients associated with Nondirected Search Usage are insignificant in both columns. The results in Table 7 reinforce our confidence in the findings from our earlier cross-sectional analyses.

#### 6.2. Using Unit Sales Instead of Dollar Sales

So far, we have been using dollar sales as the measure of sales. We now use an alternative measure, unit sales, which are not confounded by variations in product prices. Because unit sales are count data, the appropriate econometric model to use is a negative binomial regression model (Hausman et al. 1984, Greene 2002):

$$f(y_i | X_i) = \frac{e^{\mu_i \Sigma y_i^{y_i - 1}}}{y_i!}, \quad y_i = 0, 1, 2, 3, \ldots, \quad (7)$$

where \(y_i\) is the number of items purchased by consumer \(i\) in April 2006; \(X_i\) is a vector of explanatory variables; \(E(y_i | X_i) = \mu_i = \exp(X_i \beta + \epsilon_i)\) is the conditional mean; and \(\epsilon_i\) is the error term that follows a log-gamma distribution, with \(\mu_i \sim \Gamma(\mu_i / \theta, \theta)\).

We analyze the relationship of technology usage with the unit sales of promoted products as well as with the unit sales of nonpromoted products. These results are reported in Table 8. They are consistent with the results in Table 3, further indicating the robustness of our results.

---

\(^{19}\)One could also argue that consumers gain considerable information when discovering nonpromoted products by using the recommendation system, whereas they gain little information when finding promoted products through the use of this system. This disparity in information gains could also drive the skew toward nonpromoted products.

---

### Table 7

<table>
<thead>
<tr>
<th>Variable</th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Directed Search Usage</td>
<td>328.162(^*)</td>
<td>-91.161(^*)</td>
</tr>
<tr>
<td>Nondirected Search Usage</td>
<td>-115.309</td>
<td>16.044</td>
</tr>
<tr>
<td>Recommendation System Usage</td>
<td>99.147(^*)</td>
<td>23.871</td>
</tr>
<tr>
<td>Number of observations</td>
<td>2,660</td>
<td>2,660</td>
</tr>
<tr>
<td>Number of panels</td>
<td>1,192</td>
<td>1,192</td>
</tr>
</tbody>
</table>

Note: Robust standard errors are in parentheses.  
\(^* p < 0.05; \quad ^{*} p < 0.01\).
Table 8  The Effect of Technology Usage, Using Unit Sales

<table>
<thead>
<tr>
<th>Variable</th>
<th>(1) Sales of promoted products</th>
<th>(2) Sales of nonpromoted products</th>
</tr>
</thead>
<tbody>
<tr>
<td>Directed Search Usage</td>
<td>5.325** (0.415)</td>
<td>−17.989** (1.565)</td>
</tr>
<tr>
<td>Nondirected Search Usage</td>
<td>1.353 (0.939)</td>
<td>−2.793 (2.234)</td>
</tr>
<tr>
<td>Recommendation System Usage</td>
<td>1.662** (0.268)</td>
<td>3.267** (0.488)</td>
</tr>
<tr>
<td>Days Since Last Purchase</td>
<td>0.010 (0.009)</td>
<td>−0.035 (0.016)</td>
</tr>
<tr>
<td>Historical Total Purchases</td>
<td>0.026 (0.018)</td>
<td>0.141** (0.035)</td>
</tr>
<tr>
<td>Intercept</td>
<td>0.683** (0.062)</td>
<td>−0.424** (0.119)</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>−16,403.9</td>
<td>−8,228.9</td>
</tr>
<tr>
<td>Sample size</td>
<td>8,199</td>
<td>8,199</td>
</tr>
</tbody>
</table>

Note. Robust standard errors are in parentheses.
∗p < 0.05; ∗∗p < 0.01.

Table 9  The Effect of Technology Usage, Controlling for Consumer Heterogeneity

<table>
<thead>
<tr>
<th>Variable</th>
<th>(1) Sales of promoted products</th>
<th>(2) Sales of nonpromoted products</th>
</tr>
</thead>
<tbody>
<tr>
<td>Directed Search Usage</td>
<td>422.366** (62.104)</td>
<td>−127.944** (15.570)</td>
</tr>
<tr>
<td>Nondirected Search Usage</td>
<td>132.037 (149.523)</td>
<td>−18.413 (63.221)</td>
</tr>
<tr>
<td>Recommendation System Usage</td>
<td>200.707** (50.106)</td>
<td>67.169** (16.476)</td>
</tr>
<tr>
<td>Days Since Last Purchase</td>
<td>5.008** (1.281)</td>
<td>0.129 (0.446)</td>
</tr>
<tr>
<td>Historical Total Purchases</td>
<td>6.188** (0.864)</td>
<td>1.112** (0.267)</td>
</tr>
<tr>
<td>Intercept</td>
<td>20.054* (9.483)</td>
<td>7.439* (3.157)</td>
</tr>
<tr>
<td>R²</td>
<td>0.038</td>
<td>0.030</td>
</tr>
<tr>
<td>Sample size</td>
<td>3,569</td>
<td>3,569</td>
</tr>
</tbody>
</table>

Note. Robust standard errors are in parentheses.
∗p < 0.05; ∗∗p < 0.01.

6.3. Considering Potential Self-Selection Effect Arising from Directed Search Usage

One may argue that the consumers who use the directed search technology may be self-selected to be different from those who do not use such technology. For instance, consumers’ directed search usage may indicate their strong interest in making purchases (Moe 2003). Such a self-selection effect could have confounded our results. A standard way to control for self-selection effects is the propensity score matching method suggested by Rosenbaum and Rubin (1983). The idea is to obtain a sample of the control group (i.e., consumers who do not use directed search in our case) that matches the treatment group (i.e., consumers who use directed search) on many observable dimensions. Such a sample matching approach drastically reduces the difference between the control group and the treatment group, thus controlling for the self-selection effect.

We create a dummy variable indicating whether a consumer belongs to the treatment group (i.e., consumers with positive Directed Search Usage) and estimate a logit model with this dummy variable as the dependent variable and a set of observable variables as independent variables. The independent variables include the ones that measure consumers’ historical purchases ($T_{1i}$, $T_{2i}$); those that measure their income and education; and also those that measure their browsing patterns (number and duration of sessions, number of clicks, and number of category pages viewed).

We thank both the associate editor and the department editor for suggesting the propensity score matching method. We have used the STATA PSMATCH2 module by Leuven and Sianesi (2009).

Next, we perform a nearest-neighbor matching algorithm based on the propensity scores calculated in the previous step to create a matched control group. We find that before matching, the control group is significantly different from the treatment group on dimensions such as historical purchasing measures, the number of clicks, and the number of category pages. However, after matching, the difference between the matched control group and the treatment group is insignificant on all the dimensions. Finally, we repeat our analyses in §5.1 by running regressions on a “matched sample” that includes the matched control group and the original treatment group, and we report the results in Table 9. These results are very similar to those in Table 3, indicating that our main results are robust to considering the self-selection effect discussed here. In addition, one could add the propensity scores as a control variable to the original model in Equation (1). We find that our main results in Table 3 remain qualitatively the same when this new control variable is included.

20 Using other matching algorithms and using a probit model instead of a logit model do not change our results.
21 We note that Heckman’s selection model is widely applied to nonrandomly censored or selected samples (Heckman 1979). Because both the dependent and independent variables in our data are always observable and the power of our analyses comes from utilizing the whole sample, Heckman’s model is not ideal for our data. Nevertheless, we have tried this model, too. Although it performs artificial censoring of our data, the results remain by and large the same.
7. Concluding Remarks

Virtually all Internet companies provide advanced technological features such as search functions and recommendation systems on their websites, and there is plenty of anecdotal evidence suggesting that these technologies can significantly enhance consumers’ shopping experience and influence companies’ sales. Surprisingly, however, there has been very little empirical research on consumers’ usage of such technologies and the effect of consumers’ technology usage on online sales. This paper attempts to fill this void in the literature by analyzing a unique and rich data set collected from a retailing company. Our analyses show that consumers’ IT usage has a significant impact on the sales to them but that this impact varies for different technologies. More specifically, the consumers who use the retailer’s search function to perform more directed search tend to purchase more promoted products and less nonpromoted products, whereas the consumers who utilize the retailer’s recommendation system more tend to purchase more promoted products as well as nonpromoted products. Furthermore, consumers’ usage of the retailer’s search function to perform nondirected search has an insignificant effect on the sales to them.

The findings in this paper are quite robust and have survived a wide range of robustness checks. The data used in our paper come from a large Internet retailer of women’s clothing. Our concentrated data collection from one firm allows us to study the effect of technology usage on consumers’ purchasing patterns. It also enables us to measure consumers’ historical purchasing patterns that can act as effective controls for consumer heterogeneity, eliminating consumer heterogeneity as a confounding factor. Our results remain qualitatively unchanged when we consider the possibility that technology usage and sales are simultaneously determined and estimate a system of simultaneous equations. We also consider the possibility that the consumers who use technologies are self-selected to be different from those who do not use technologies, further verifying the robustness of our results.

Our findings are likely to have broad applicability. First, the search and recommendation technologies we study are widely adopted by most Internet websites selling products in the apparel category as well as in many other categories. We acknowledge, however, that the size of the effect may vary for other companies depending on the quality of technology implementations at those companies. Second, the retailing company we study employs operations and marketing strategies that are widely used in the retailing industry. Most retailers promote their products in different channels—catalogs, flyers, e-mails, TV commercials, etc.—and consumers are likely to have prior knowledge on some products but not on others before visiting these retailers’ websites. Thus, we expect our findings that the effect of technology usage varies across different products to hold in scenarios where consumers have varying levels of prior knowledge on different products. Finally, the empirical models we develop in this paper can be readily applied to quantify the effects of any IT-enabled tools that alter consumers’ information search behavior.

We also expect our results to be qualitatively similar for other product categories although, once again, the size of the effect may vary. Products in most categories have SKUs, numbers, or names that act as unique identifiers. As long as consumers can link product knowledge they have to these identifiers, it is likely that their use of directed search will improve the efficiency with which they recall prior knowledge, which, in turn, is likely to have an effect on their external information search as well as on the sales to them. In addition, in any product category with a large number of SKUs (e.g., books, music, DVDs), searches with generic and nonspecific keywords are likely to produce an excessively long list of products, as they often do in the clothing category. Hence, in these product categories, the use of nondirected search (particularly when a ranking of search results is unavailable) is likely to remain ineffective in facilitating consumers’ external information search and ultimately in affecting the sales to these consumers. Finally, we expect the recommendation system to have a relatively larger impact in product categories with many product SKUs than in product categories with few SKUs. In categories with many SKUs, consumers are likely to lack prior knowledge about a large proportion of the overall product assortment and therefore may find the website’s recommendations more beneficial in facilitating their external information search.

Understanding the impact of IT on sales has important managerial implications. First, the results in this paper show that consumers’ technology usage is generally associated with a higher level of sales to them, providing economic justification for firms’ investments in advanced technological features such as search functions and recommendation systems. We should, however, caution that our study uses archival data rather than experimental data. We have used several econometric techniques to check for simultaneity and self-selection, as well as other plausible effects, and found our results to be robust. Future research is needed to verify the causal effect of IT usage on online sales, and such research may require lab or field experiments that directly manipulate consumers’ access to various website technologies. Second, our results suggest that firms should strategically encourage consumers to use technologies...
available on their websites. There is an extensive literature on user’s acceptance of IT (e.g., Davis 1989), and the knowledge from that literature could be applied to improve consumers’ acceptance of website technologies. Furthermore, our results demonstrate that consumers’ use of directed search skews their purchases toward promoted products, accentuating the effect of firms’ marketing promotions. On the contrary, as shown in §5.4, consumers’ use of recommendation systems skews their purchases away from promoted products, mitigating the effect of marketing promotions (although these systems enhance the overall sales of both promoted and nonpromoted products). Third, our results also suggest that firms should invest in IT because even the mere presence of these technologies may have a significant positive impact on sales. For example, the presence of a recommendation system increased the sales by more than 5.5% for the retailer used in this study. All in all, these results imply that firms should seriously consider the affects of IT when planning marketing promotions.

This paper demonstrates that information technologies can change the way consumers seek information from internal and external sources, affecting the sales to online consumers and their purchasing patterns. As Internet companies continue to develop ever-more-sophisticated search and recommendation technologies, and as consumers continue to increase their use of these IT-enabled tools, the importance of the effects of IT usage on online sales, like the ones identified and examined in this paper, is only going to grow.

Acknowledgments

The authors thank Erik Brynjolfsson, Jerry Hausman, Barrie Nault, Dabao Zhang, and the seminar participants at Purdue University and the University of Maryland as well as the participants of the 2007 Workshop on Information Systems and Economics (WISE), the 2008 Big Ten Information Systems Symposium, and the 2008 INFORMS Annual Meeting for valuable comments on this paper. The authors also thank the department editor, the associate editor, and two anonymous reviewers for their guidance and helpful suggestions. The authors are listed in alphabetical order.

Appendix. Correlation Matrix

<table>
<thead>
<tr>
<th></th>
<th>Sales of Promoted Products</th>
<th>Sales of Nonpromoted Products</th>
<th>Overall Sales</th>
<th>Directed Search Usage</th>
<th>Nondirected Search Usage</th>
<th>Recommendation System Usage</th>
<th>Days Since Last Purchase</th>
<th>Historical Total Purchases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sales of Promoted Products</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sales of Nonpromoted Products</td>
<td>0.035</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall Sales</td>
<td>0.945</td>
<td>0.361</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Directed Search Usage</td>
<td>0.123</td>
<td>-0.111</td>
<td>0.078</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nondirected Search Usage</td>
<td>0.028</td>
<td>-0.028</td>
<td>0.017</td>
<td>0.096</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recommendation System Usage</td>
<td>0.062</td>
<td>0.092</td>
<td>0.088</td>
<td>-0.103</td>
<td>-0.006</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Days Since Last Purchase</td>
<td>0.008</td>
<td>-0.045</td>
<td>-0.007</td>
<td>-0.012</td>
<td>0.018</td>
<td>0.022</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>Historical Total Purchases</td>
<td>0.094</td>
<td>0.084</td>
<td>0.115</td>
<td>0.035</td>
<td>-0.015</td>
<td>-0.027</td>
<td>-0.603</td>
<td>1.000</td>
</tr>
</tbody>
</table>

References


De, Hu, and Rahman: Technology Usage and Online Sales: An Empirical Study
Management Science 56(11), pp. 1930–1945, © 2010 INFORMS


Leuven, E., B. Sianesi. 2009. PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. Statistical software component S432001. Boston College Department of Economics, Chestnut Hill, MA.


