Online Retailer Service: IT Features and Price Dispersion

Neveen Farag Awad

University of Michigan Business School
Ann Arbor, MI 48105
nfarag@umich.edu

M. S. Krishnan

Michael R. and Mary Kay Hallman e-Business Fellow
Associate Professor and Area Chair of Computer Information Systems
University of Michigan Business School
Ann Arbor, MI 48105
mskrish@umich.edu

November 20, 2003

Financial support for this study was provided in part by the Michael R. and Mary Kay Hallman fellowship, the Marcy Maguire fellowship, the Information Systems Executive Forum at the University of Michigan Business School, and the Summer Research Opportunity Program at the University of Michigan.

Online Retailer Service: IT Features and Price Dispersion

Abstract

It has been theorized that the online environment has lower search costs due to the ease of comparing prices across various retailers. However, recent empirical evidence has shown that despite potential lower search costs, prices for a homogeneous good do not converge; instead, price dispersion, the distribution of prices across vendors, is greater online than offline (Smith et al. 2000). It has been argued that access to information facilitated by e-business technology has shifted the consumer's role from a passive receiver of products to that of an active member in defining and creating value. Consequently, in order to attract new customers and improve customer loyalty and profitability, firms are utilizing consumer-provided information to more effectively personalize the user's online experience. In this paper, we examine certain features of online service. In particular, we review personalization factors and their association with a firm's ability to charge a price premium. Our empirical examination includes analysis across 79 online retailers for 17 products including books, compact discs (CDs), and personal digital assistants (PDAs). Our results indicate that the implementation of certain user-interaction features and information are positively associated with personalization. An interesting finding is that while product information is negatively associated with an online retailer's price premium, increased personalization is positively associated with price premiums for books and PDAs.

Introduction

Different prices charged for the same good by different sellers is referred to as "price dispersion". Price dispersion is important to study as it can significantly affect consumer and producer behavior; indeed, it may imply potential gains to firms through decreased consumer search. The existence of price dispersion has been widely studied across various environments including environments conducive to perfect competition (Pratt et al. 1979, Dahlby and West, 1986, Sorenson 2000). Studies have shown that in frictionless e-commerce, prices for identical products sold by different online retailers are expected to converge (Bakos 1997). However, recent empirical studies show that such a scenario does not exist. Such evidence of price dispersion in online markets may challenge the hypothesis that e-commerce leads to increased market efficiency through reduced search costs. Indeed, while several studies have shown that price dispersion exists online (Brynjolfsson and Smith 2000; Clemons, Hann. and Hitt, 2002), few have explored the information technology (IT) features that a firm may implement towards better online service; such features would consequently encourage less price comparison by consumers. Hence, we examine various factors that may be associated with online search costs, and their subsequent effect on price. By examining these online features, we seek to provide insight into how a firm can better design their website to enhance customer value and therefore charge price premiums. While price premium is not necessarily a signal of financial success, it has been asserted as a sign of market power (Gazzale and Mackie-Mason 2000).

Incomplete information is the most often cited theoretical explanation of price dispersion (Stigler 1961). Subsequently, researchers have modeled price dispersion as an equilibrium outcome by incorporating the cost of locating information. For example, price dispersion occurs when consumers find that the effort involved in finding the lowest-priced product outweighs the

cost savings. It follows, therefore, that by decreasing search costs, or the costs of locating price information, price dispersion should be reduced. In this paper, we examine various factors that potentially affect search costs and the corresponding effect on price premiums.

While incomplete information is one theory for the existence of price dispersion, other explanations have been offered: Fishman (1992) suggested that staggered price setting due to menu costs could be a source of price dispersion, as some sellers may not be able to adjust prices as quickly as others. Dana (2001) suggested that demand uncertainty, costly capacity coupled with presetting of price, may drive price dispersion. It is worth noting that Dana's model is based on the assumption that prices are rigid, which is in contrast to the online environment. Nonetheless, following Dana's explanation, as a market becomes more competitive, prices become more dispersed. Price discrimination is a third possible explanation of price dispersion: Price discrimination is the practice of charging different prices to different consumers according to their willingness to pay; such a trend is shown to exist in the online travel agent market (Clemons, Hann. and Hitt 2002), as well as the Italian grocery market (Giuletti, 1999). Lastly, differential services, or unmeasured heterogeneity in seller attributes, may lead to price dispersion in markets for homogenous goods (Brynjolfsson and Smith 2000; Sorenson 2000). To our knowledge, this current research is the first to quantify various online retailer website features assess their impact on price dispersion.

While several theories of price dispersion exist, empirical evidence to support our understanding of what drives online price dispersion is scarce. Prior empirical studies focus on only a few classically measured factors, such as product heterogeneity (Clemons, Hann, and Hitt 2002; Dahlby and West 1986; Erevelles, Roland, and Srinivasan 2001; Giuiletti 1999; Pratt, Wise and Zeckhauser 1979; Smith et al. 2000; Sorenson 2000). In this study, we extend prior

work, which identified the presence of online price dispersion and suggested possible determinants related to seller-side attributes (Brynjolfsson and Smith 2000; Smith et al. 2000). To date, how companies manage price dispersion remains unclear. Therefore, our research tests the suggested theory that online firms can influence search costs and consequently induce price dispersion through various information systems factors.

In the next section, we review and discuss prior literature and theoretical foundations relevant to our work. In section three, we present the research model framework and discuss the hypotheses. In section four, we explain the data and measurement. In section five, we present the analysis and results. In section six, we discuss the results and their corresponding implications. We then conclude the paper and offer direction for future research.

2) Prior Literature

The increased access to information facilitated by Internet growth has received the attention of researchers. It has been argued that this increased access to a large number of consumers may lower online search costs (Bakos 1997). This is because while conventional retailers are usually unwilling to provide many price quotes when contacted, prices are available online through shop-bots or price comparison web sites, such as DealTime.com, BizRate.com, and MySimon.com (Sorenson 2000). Hence, the growth of online retail has provided a great resource for studying the existence of price dispersion.

Many researchers have studied consumer search costs in commodity markets, where all sellers offer identical products such as books, compact discs (CDs), computer hardware, computer software, personal digital assistants (PDAs), and video games. For example, Brynjolfsson and Smith (2000) and Clay, Krishnan, and Wolff (2001) empirically illustrate that price dispersion exists online in the book industry. While these studies have enriched our

knowledge regarding the degree of price dispersion and possible explanations for price dispersion, the focus of most of these studies have been on a given product category, such as a books or CDs. Notwithstanding this, it is recognized that determinants of price dispersion may differ among various product types (Brynjolfsson and Smith 2000, Clay, Krishnan, and Wolff 2001). In this study we examine price dispersion in three product categories: books, CDs, and PDAs. In choosing to focus on these particular product categories, our aim is to explore the association of online service factors to price dispersion as it compares across products of varying characteristics that are consistent with prior theory (Peterson et al. 1997). Such variations include frequency of purchase, price, degree of heterogeneity in consumer taste, and variance in consumer-profiling with regard to their sophistication and knowledge about the relevant product.

Previous online price dispersion studies often focus on higher-level determinants, such as brand recognition or a consumer's previous experience with an online retailer (Brynjolfsson and Smith 2000). The need to understand lower-level firm-specific features that may enhance service quality and their association with price has been identified in prior research (Sorenson 2000). In this study, we examine some of these factors, i.e. features offered by the firm to enable consumer interaction with websites, as well as online personalization of information. When we speak of personalization, we speak in terms of the definition provided by Nunes and Kambil (2001): "The process of gathering information explicitly or implicitly about a customer, which enables the firm to target products or recommendations that best match the consumer tastes." Previous studies have analytically examined the effect of online personalization on consumer preference (Farag and Van Alstyne 2000). However, none of the prior studies of online price dispersion have attempted to empirically measure IT features of a website. This research is the first to introduce a set of metrics for assessing some of these unmeasured online

service factors related to website features. Specifically, we test the hypothesis that the use of IT website features (such as personalization, user interaction, and product information offerings), are associated with a firm's price premium.

Theoretical Foundations

Although e-business has shown some promise, the post "dot-com-boom" era finds firms struggling to find the right combination of attributes of e-business infrastructure in order to succeed. What constitutes a successful e-business infrastructure is still widely debated; yet, the importance of such an infrastructure on the future of businesses remains unquestioned (Weill and Vitale, 2001). Through case studies, Weill and Vitale (2001) identified the collection and use of a reliable customer database as one of the most important infrastructure services for understanding and responding to the customer. In this paper, we explore how firms can use their customer information to better understand customer needs. By tailoring the presentation of product and price information on their websites, firms can personalize the online shopping experience for each customer and earn premium prices for their goods.

Online consumers face two types of search costs: 1) the costs of obtaining price information, or the cost of price comparison *across* web sites, and 2) the cost of obtaining product-characteristic information, or the cost of finding a product that fits the consumer's preferences *within* a website (Bakos 1997). The exact relationship of price and product search costs to equilibrium prices is still unclear (Harrington 2001). However, firms can potentially influence these search costs through their online features. In online purchasing, for example, many firms do not show the full price of a product, including shipping and taxes, until the credit card information is entered. In such a case, the cost of obtaining price information is high. We believe that underlying IT

system design can influence search costs associated with price and product information. For example, the use of collaborative filtering is an attempt to decrease product information search cost by personalizing the product information presented to the user. Such an attempt is consistent with known information of the product and user expectations (Shardanand and Maes 1995). Firms can implement various service offerings via their IT system in an attempt to alter price search costs and product search costs. These service offerings include product information, personalization, and user interaction. Price premium is the primary dependent variable of interest; indeed, it has been used as a proxy for market power in past studies (Diamond, 1985; Stiglitz, 1989). The rationale surrounding the use of price as a proxy is that if a consumer values an e-business offering at a price level such that the firm does not need to discount its prices, such value potentially signals a successful business model. However, we fully recognize that price, in and of itself, is not a signal of financial success. However, the spread of prices across homogenous product offerings is a sign of market inefficiency; it is the use of IT as the source of this inefficiency that we investigate in this paper. Table 1 summarizes the factors examined in our study along with descriptions and variable names. We address each of these factors individually in our hypothesis section.

_____See Table 1_____

3) Research Model and Hypotheses

The first stage of the model investigates factors that enhance personalization. In order to lead the competition in electronic marketplaces, firms may attempt to differentiate themselves through methods used to present product information and price information. For example, many online retailers make all of their product information readily available to consumers. In the case of book online retailers, for instance, it is common to find pages of information about a certain

book on an online retailer's website. Product information may be more useful, however, when firms utilize IT to capitalize on personalizing the shopping experience. Lynch and Ariely (2000) found that increasing the ease of cross-firm price comparison only increased consumer price sensitivity for products (in their case wines) common to many stores. However, for unique products, cross-store comparison had no effect on price sensitivity. IT infrastructures can convert the product information into useful information for consumers. Personalization features of online retailers typically work through mechanisms such as collaborative filtering (Shardanand and Maes 1995). These features are often built on pattern recognition methods used to identify and extrapolate from trends in customers' choices, information usage, and/or demographics (Nunes and Kambil 2001). Thus, we expect that the more product information and customer information an e-business has available in their infrastructure, the more personalization they may offer their consumers. In fact, company case studies have suggested that as a firm's consumer database grows, the companies move from rules-based data mining to collaborative filtering systems in order to better deal with increased information (Kroll 2001). Thus, Hypothesis 1 follows:

Hypothesis 1(H1): A greater amount of **product information** is associated with a greater amount of personalization.

It has been argued that customer-perceived value innovation is achieved through improving performance on customer-identified objectives (Kim and Mauborne, 1997). By encouraging user interaction through rating systems and comment sections, firms are able to facilitate a dialogue with their customers. This dialogue provides customers a sense of involvement and interaction in their shopping experience, and allows firms to monitor characteristics and features that consumers deem valuable. Customer information is collected by

online retailers through user interaction. When customers store items for future purchase in their shopping cart firms are able to deduce consumer preferences. Personalization technologies need customer information, along with product information, in order to make useful recommendations. Thus, increased user interaction should give a strong foundation to collaborative filtering mechanisms, allowing for greater personalization. Hypothesis 2 follows:

Hypothesis 2(H2): A greater amount of **user interaction** is associated with a greater amount of personalization.

In the second stage, we examine how three factors of a firm's website, namely information, user interaction, and personalization, are associated with an increase in **price premium**. We first examine the dependent variable of the first stage, personalization. Personalization enhances value to consumers through improved information accuracy and relevance for a given context. A possible explanation of the effect of personalization on price is that it decreases the search cost of product information; in other words, we stipulate that when buyers are provided with relatively accurate information by an e-business regarding which product best fits their preferences, their incentive to search across other firms decreases. As a consequence, personalization can lead to a firm's ability to charge a price premium for products. Amazon.com is a classic example of a firm that attempts to use personalization to drive down product search costs. By gathering and using consumer data effectively, Amazon.com and other firms attempt to present the products that most closely match customer needs. Thus, personalization may decrease consumer search cost for product information, and therefore increase price premium. Hypothesis 3 follows:

Hypothesis 3(H3): A greater amount of **personalization** is associated with a larger price premium.

Next we examine the direct effect of product information on price premium. We believe that product information presented in a firm's website can have both *direct* and *indirect* effects through personalization on price premium charged. It is likely that all of the product information presented may not be relevant to what the user requests and may even lead to information overload. Hence, increased dimensionality of product information presented may lead to worse consumer decision-making and also increase consumer search cost for product or price information. However, we expect that the *indirect* effect of product information on price premium through personalization will be positive: more product information will enable firms to better personalize the contents for a given user and, as a result, create a capacity for charging higher price premium as previously discussed. The key idea between the differences in the direct and indirect effects of product information on price premium is that while people are not qualified to deal with large amounts of product information, properly designed information systems may be able to utilize such information to personalize the user's content.

It has been argued in prior literature that the nature of price and product information can significantly affect a firm's ability to charge premium prices for its products or services (Bakos 1997). Increased consumer search costs for *pricing* information has been shown to enable firms to charge price premiums due to greater market power (Salop and Stiglitz, 1977). The opposite is true for product-information, whereby easily accessible product information allows for a more accurate assessment of what good best fits the needs of the consumers. Therefore, *decreased* consumer search costs for product information in electronic marketplaces have been shown to increase market power, leading to increased price premiums (Bakos 1997). Consequently we suspect that an increase in the amount of product information provided by online retailers

increases product information search costs and therefore decreases price premiums. Hypothesis 4 follows:

Hypothesis 4(H4): A greater amount of **product information** is associated with a smaller price premium.

Lastly, we examine user interaction. User interaction enables consumers to reveal their preferences by exercising some level of control over the content of the site. Firms can then use IT to target consumers with coupons and special offers aligned with their revealed preferences. Such targeting can potentially increase customer loyalty. In addition, options to save preferences may decrease consumer product information search costs, as products can be stored for future purchase. It is worth noting that too much interaction can also annoy consumers and lead them to switch to another online retailer. In assessing user interaction, we focused on transaction-based items, including shopping cart and order history storage. In focusing on transaction based interaction items, we are attempting to focus on items central to the purchasing experience, rather than tangential interaction items that may increase, rather than decrease product information search costs. Thus, the potential decrease of product information search costs within a site brought about by increase user interaction may result in an increase in price premiums. Hypothesis 5 follows:

Hypothesis 5(H5): A greater amount of **user interaction** is associated with a larger price premium.

Product Category Differences

As noted earlier, we expect the effect that online service factors have on the online shopping experience, and therefore the firm's posted price, to vary across product categories. Peterson et al. (1997) argue that the suitability for firms to use the Internet depends on the

products and services they sell. Thus, Peterson et al. classify products and services along three dimensions: 1) Cost and frequency of purchase; 2) Value proposition; and 3) Degree of Differentiation. The scale of each respective dimension is as follows: 1) low-cost, frequently purchased goods (milk) to high-cost, infrequently purchased goods (stereos, palm pilots); 2) Tangible or Intangible; and 3) High differentiation potential (wines, online newspapers) to Low differentiation potential (stock market quotes, diamond of known color, weight, and clarity). We build on this categorization in choosing our three products, and keep fixed the second two dimensions of Peterson et al.'s classification scheme. Thus, we chose three tangible products that across the entire category had a large potential for differentiation. The three product categories we chose were: books, CDs, and PDAs. We controlled for degree of differentiation across all the three categories by gathering posted price data for specific books (e.g. Think and Grow Rich is the same book regardless of where it is purchase from). Thus, while the degree for differentiation is relatively high across the product categories, in that there are terrible books, and there are great books, similar with CDs and PDAs; this can be controlled for using specific titles or model numbers.

The main difference of PDAs from Books and CDs is that PDAs are higher priced and less frequently purchased goods relative to books and CDs. In addition, PDAs are more complex products that may require user understanding of the various features as well as details in compatibility with various computer applications. Books and CDs, however, fall into the same product categories according to Peterson et al.'s classification, as they are both tangible, relatively low cost, high frequency goods that across the product category have high differentiation. However, that there is a sub-element of product differentiation that comes into play in product classification; that sub-element element is: uncertainty in consumer expectations.

Books are classic commodity products, and illustrate why previous studies have focused on the book industry (e.g. Brynjolfsson and Smith 2000; Clay, Krishnan, and Wolff 2001). However, music products, (CDs in our study) are more like experience goods, as consumers often have limited knowledge of an entire CD's content before purchase, (e.g. Gu and Hitt, 2001). Thus, once a specific book or CD is chosen, the perceived quality of the product can have different levels of dispersion. We expect, therefore, that the uncertainty of consumer expectations is higher in the case of music (CDs) than books or PDAs, due to music being more of an experience good. In summary, the categorization provided by Peterson et. al has helped us better classify the product categories that we tested.

We tested our theory using review data of 10 books, 10 PDAs, and 10 CDs from Amazon.com. For each product we collected the 60 most recent reviews and assessed the variance across a five-point scale. The average variance across the 10 books was 0.48. Similarly, the average variance across the 10 PDAs was 0.41. However, the average variance across the 10 CDs was 1.54. A t-test confirms that the sample of CD variances was significantly different from the sample of variances from Books as well as PDAs at the $p<0.001$ level. In addition, previous literature has suggested that preferences for music are less homogenous across a random group of consumers than for books (Kroll 2001). These empirical results support our theory that there is a different degree of uncertainty in consumer expectations across the book and CD category; this may be due to the more experience nature of CD products. Thus the difference in uncertainty of consumer expectation in the CD category may pose a challenge to classic personalization technologies, such as collaborative filtering, which try to identify common threads from aggregated consumer preferences. There may be a need for

personalization technologies to be geared up for experience goods, such as music, to address the greater variance in consumer preferences. Thus, Hypothesis 6 follows:

Hypothesis 6(H6): In more experienced based goods (such as CDs in our sample), a greater amount of **personalization** is associated with smaller price premium.

In summary, we believe that by including books, CDs, and PDAs in our study, we are able to "tease out" the potential structural differences, (if any exist), in the determinants of price premium across product categories. In the following section, we describe our research setting to test the model.

4) Data and Measurement

In this section, we describe the research site and data collection methods to test the model. A schematic diagram of the conceptual elements of our research model is shown in Figure 1. The model addresses the research questions related to the direct and indirect factors affecting online price premium. Past researchers have approached the study of macro-variables, such as price, through the use of an intermediary variable in a two-stage model (e.g. Barua, Kriebel, and Mukhopadhyay, 1995). We follow this approach, and utilize personalization as the intermediary variable to understand price premium determinants. Personalization was hypothesized to be the mediating variable, as it takes raw product and consumer information and converts it to targeted, user specific information steps. We test for Personalization as the mediating variable using a three-step mediation test suggested by Baron and Kenny (1986). Accordingly, a variable may be considered a *mediator* to the extent to which it carries the influence of a given independent variable (IV) to a given dependent variable (DV). In our case the given independent variables is information, and the given dependent variables is

personalization. Thus, in this paper, mediation can be supported if the following holds true: 1) the *information* variable significantly affects *personalization;* 2) the *information* variable significantly affects the *price premium* variable in the absence of the mediator; 3) the *personalization* variable has a significant unique effect on *price;* and 4) the effect of *information* on *price* shrinks upon the addition of *personalization* to the model. The *Sobel Test* has been popularized by Baron and Kenny (1986) as means for testing whether a mediator (personalization) carries the influence of an independent variable (information) to a dependent variable (price premium).

The sobel test statistic produced with our data sample is 1.97, which is significant at the $p<0.05$ level. Thus, the indirect effect of information on price premium via the information variable as the mediator is significantly different from zero.

Research Site and Data Collection

Our research setting spanned across 27 online book retailers, 27 CD retailers, and 25 PDA retailers. Data was collected across 6 books, 5 CDS and 5 PDAs. In our data collection process, we collected both cross-sectional and longitudinal data from a variety of sources. Control variable data was collected from Google.com and BizRate.com. Website features counts were collected directly from the online retailer websites. This study uses a time-slice of the cross-sectional data to explore the influence factors on price levels as well as the factors that indirectly affect pricing through personalization.

Exploratory Research

With the help of groups of online consumers, we first identified characteristics of the online shopping experience that are perceived as valuable by consumers. Our exploratory

research involved several focus groups with MBA students who shopped online. A major purpose of these groups was to identify factors that customers deemed to be of value.

The focus groups revealed that online retail customers derive value from information provided in various ways, whether it be product information, personalized information, or stored, interactive, transaction information. At the time of the study, online personalization, user interaction, and product information were not yet tested in the literature; consequently, robust scales did not exist to guide the measurement process. Therefore, we created a set of attribute measures to assess these variables; these measures are grounded in theory introduced in prior literature. First, we based personalization feature counts on psychology literature demonstrating that when a person perceives that another entity is familiar with them, the person increases compliance behavior (Berscheid and Walster, 1978) as well as reciprocal positive feelings (Curtis and Miller 1986; Drachman, de Carufel, and Insko 1978; Jacobs, Berscheid and Walster, 1971; Regan 1976). Second, we based user interaction measurement on media richness theory literature: media richness refers to a medium's relative ability to communicate a message (Daft and Lengel 1986). According to media richness theory, online interaction should provide users richer capabilities than classic catalogues with pictures and text (Palmer 1997). Thus, we attempted to assess the richness of the medium through interaction functionality provided for the consumer. Third, we based product information measurements on literature that shows that locating high-quality information within the computer-mediated context is important (Hoffman et al. 1995) and providing high-quality information directly related to a product's salient attributes enhances consumer response (Alba et al. 1997).

Methodological Steps Taken to Assure Rigorous Results

The research team was composed of five student members who collected field data across 79 online retailers. Each research member was given the same document that described the features to be assessed at each website. The aim was to make the data collection as objective as possible, therefore all of the measurements were either binary (access to book excerpt? (yes/no)), or were objective counts of website features (how many times does the user name appear on the home page after having logged in?). For each online retailer, feature assessments were gathered across different book, CD, or PDA offerings. In addition to the measurements being as objective as possible, several other data collection parameters were controlled to assure rigorous results. These environmental controls include:

- *Location*—All research members sat in the same computer lab, and used the same computer model to collect the data.
- *Time*—All research members collected data at the same time and days each week, four times a week over a two-month period, from early June 2001 through July 2001
- *Who measured*—the research team members stayed consistent.
- *Websites visited*—all research team members assessed all the websites, in the same order as the other members. The order of assessment was randomized each day, such that a different company was the starting point each day. However, each research team member assessed the same company at the same time.
- *Indications of data quality* – Weekly measures of interrater reliability were assessed, to confirm that research team members were consistent in the measurements. The aggregate interrater reliability metrics are included below.

Choice of Online retailers

We chose the top 30 rated online bookstores, CD retailers, and PDA retailers according to BizRate.com; BizRate.com surveys online retailer customers and asks them to evaluate online retailer services. BizRate.com maintains online ratings by allowing consumers to complete a satisfaction survey upon purchasing a product online. Numerous price comparison portals, as well as online retailer sites, refer to BizRate.com's ratings; thus, we use the mean rating in

several categories as popular indexes of online retailer heterogeneity factors. The running means of ten aspects of online retailer service, evaluated with a ten-point scale, are posted on BizRate.com's website. From the overall customer satisfaction rating, the top 30 sites in customer satisfaction were chosen.

Choice of Products

In order to avoid any bias from product choice, we selected and controlled for books, CDs, and PDAs from various topic areas. The range of books, CDs, and PDAs were used to separate out any product-specific bias effects. In the case of books, for example, we used titles that included a New York Times Best-Seller, a best-selling Java development book, a high-selling business book, and a number of less popular cooking, home-improvement, and philosophy texts. We chose such a range to avoid special pricing that occurs across certain classifications, for example, the New York Times Best Seller list. The five CDs ranged from classical to top-40 and from rock to jazz. The PDA's ranged across different manufacturers, from the least expensive to the most expensive models. In addition, to assess whether companies subsidized one product, and charged a premium for others, we also randomized bundles of products, and used the price premium on the bundle as a dependent variable.

Pilot Study

An initial pilot test was done where each site was rated twice over the course of one week in order to clarify ambiguities and assure consistency across the research team. The point of using items based on feature counts was to create a set of objective measures. For example, there should be no discrepancy in the number of times a name appears on the homepage after a user has logged in. By comparing the variance in response, the pilot test was designed to isolate counts that were treated differently across research team members or were otherwise unclear. To

ensure that the items measured the appropriate constructs, the construct validity of each item was examined. Kerlinger (1978) cites two methods of construct validation: (1) correlations between total scores and item scores and (2) factor analysis. The first approach assumes that the total score is valid; thus, the extent to which the item correlates with total score is indicative of construct validity for the item. Each research team member collected items that related to the construct as a whole, or feature counts. For each of the three constructs of Personalization, Product information, and User Interaction, a group of students rated each site in our sample on a 1-10 scale. The mean across these 46 students was calculated. We then used these average ratings as the total score for each of the three main constructs. We assessed the degree to which the items correlated to the total score across companies while maintaining the assumption that the overall ratings were valid. Items were eliminated if their correlation with the global item score was below 0.4. Since there is no acceptable standard cutoff, we chose a cutoff of 0.4 since it is a comparable cutoff to those used by other researchers (Doll and Torkzadeh, 1988, Ives et al. 1983). The correlations with the corrected item total ($r \geq 0.4$) were statistically significant ($p < 0.001$). Thus, the cutoff was considered high enough to ensure that the items retained were adequate measures of the e-business infrastructure constructs. The criterion enabled the researchers to reduce the 18 collected items to 13 useable items. All items were normalized to a standard 0-1 scale for factor analysis. These 13 items combined into the final 3 independently assessed e-business factors used in the study. The three factors had acceptable reliabilities (Cronbach's alpha) as shown in Table 2.

____See Table 2____

5) Analysis and Results

Factor Analysis

A total sample of 402 observations across the product types (27 book retailers, 6 books per firm, 25 PDA retailers, 5 PDAs per firm, 27 CD retailers, 5 CDs per firm) was examined. The data was examined using principal component analysis as the extraction technique. Eighteen items were factor-analyzed using a varimax rotation. The goal of the factor analysis was to extract independent attributes that aligned with theory and contribute to the online consumer experience. Following an approach used by previous researchers, we dropped items with multiple loadings across factors during the iterations of the factor analysis (Doll and Torkzadeh, 1988). Thirteen items loaded unambiguously on three factors; these results are shown in Table 3. These factors were interpreted as **personalization**, **user interaction**, and **product information**.

Reliability of the overall scale, as well as the individual factors, exceeded the accepted threshold of 0.70 (Nunnally 1967). The 13-item scale had an overall reliability of 0.7821. The reliability of each factor was: 0.9327 for Personalization, 0.8546 for User Interaction, and 0.8182 for Information.

_____See Table 3_____

Variables Definition in Empirical Model

The following variables were used in our analysis:

Personalization: As shown in Table 3, the personalization construct of an online experience is measured using five separate items:

(1) The number of times the user is referenced by name upon login

(2) The number of functionality changes before and after the user logs in

(3) The total number of links targeted to the user on the product page

(4) The number of customer reviews, and

(5) The number of customer-rating reviews

The first item, references to the user by name, included addressing the user by first name ("Joe"), last name ("Mr. Smith") and full name ("Joe Smith"). The second item, functionality changes, included new features such as a "wish list" or a gift-registry that are presented to the user upon logging in. The third item, links targeted to the user, included pointers to products bought by consumers who had bought the same product as the current user; also included were pages that were made "for you" by the firm. The fourth item includes the total number of customers reviews presented to the consumer of each product, and the fifth item is the total number of customer ratings of reviews that were presented to the customer. At the time of the study, the research team noted that the reviews and ratings presented were different to different users for the same product, which is why the reviews and review ratings were included in the personalization measure. As noted earlier, these features are grounded in psychology literature that suggests that remembering and using consumer names and likes is important (e.g. Futrell 1988, Marks 1991). Literature in applied sales has also emphasized that customer name and preference recognition should result in increased sales (Levy and Weitz 1992; Witsman 1987) and increased familiarity (Berscheid and Walster, 1978). Personalization is, in essence, automated familiarity with a consumer. Since features such as using a person's name have been shown to increase the perception of familiarity, we expect personalization will increase, as well. The items aggregated for the personalization factors are largely objective counts. For each item, several of the research team members independently measured each site, across several product offerings. The final personalization score for each firm's product offerings were obtained by averaging the independent assessments for each count. The interrater reliability index of 0.86 for this measure was well above the threshold recommended by Nunnally (1967).

User Interaction: The user interaction factor measures the degree to which the consumer is able to see and interact with information about themselves and their transactions. As shown in Table 3, the user interaction construct of an online experience is measured using a summation across two separate items:

(1) Can a customer store products in a shopping cart? (0/1)
(2) Can a customer store and access previous order information (credit card used, shipping address, products purchased)? (0/1)

In focusing on the interaction of the shopping cart and the order history, our goal was to examine transaction-based interactions. Previous case studies have suggested, but not robustly shown, that allowing consumers to interact with account information, shopping carts, and previous order history enhances online customer service (Palmer 1997). The final score of each firm's user interaction offering was a summation across the two features; thus a firm could receive 0, 1, or 2, depending upon which of the two features the firm had made available, if any. The interrater reliability index of 0.98 for this measure was also well above the threshold recommended by Nunnally (1967).

Information: The information factor measures the level of product information provided by the firm to consumers on the products in which they are interested. Previous research suggests that successful sale of products online relies on ample product information (Palmer 1997). In assessing product information, research team members counted the total number of information features that a firm had out of a list of six possible information features. Included in this list are the following:

(1) A list of 4 or more product attributes
    a. In the case of Books, this list was
        i. Dimensions

ii. Publisher
iii. ISBN
iv. Available Editions

b. In the case of CDs, this list was
i. Number of Media/Number of Discs
ii. Release Date/Street Date
iii. Label
iv. ID Number (ASIN/ CDU Part#)

c. In the case of PDAs, this list was
i. Maker (Palm, Compaq, etc.)
ii. Operating System (Palm O/S, Windows C/E)
iii. Memory
iv. Manufacturer Part Number

(2) Table of contents
(3) Book excerpt/CD sound clip (not used in PDA metric)
(4) Reviewer Information
(5) Site-provided product review
(6) Full product pricing information on product page (rather than partial price information until the product is put in the shopping cart)

The overall information score is a count of the total available information features, with the maximum possible score being six. The interrater reliability index of 0.93 for this measure was also well above the threshold recommended by Nunnally (1967).

Control Variables

Previous research has stipulated that various factors, including: for brand equity, customer satisfaction with shipping and handling, and ease of use of their shopping experience, affect price dispersion (Smith et al. 2001). Thus, we control for these various factors in our assessment of online service characteristics:

Ease of Use and Shipping and Handling: Both consumer-rated ease of website use and shipping and handling satisfaction of online stores were gathered twice a week from BizRate.com. Upon purchasing a product from an online retailer, consumers are presented a pop-up window to rate various aspects of their online shopping experiences on a scale of 1-10;

BizRate.com averages this data across all consumers that have to date rated the store. Among the factors that a consumer is asked to rate is:

1. **Ease of Use of Website** (This is asked of the consumer immediately following the purchase)
2. **Satisfaction with Shipping and Handling** (This is asked of the consumer several weeks after the purchase)

Brand: Brand is a categorical variable, extended from that used in prior studies (e.g. Brynjolfsson and Smith 2000). Table 4 outlines the coding of the categorical brand variable.

____See Table 4____

Book/CD/PDA Type: Since the ratings were done across several books, CDs, and PDAs, we controlled for which book/CD/PDA was being rated by a categorical variable to assess if firm differences also differed across book/CD/PDA type.

Number of Links: The number of external links refers to the number of other web sites that have links to the particular online retailer website. It has been argued that an increased number of external links represents greater potential traffic to a website, and therefore greater presence across the Internet (Bradlow and Schmittlein 2000). Hence, we use the external number of links as a surrogate measure of the popularity or presence of a given online retailer in our analysis. The number of links was assessed through Google.com, which will allow you to enter a Universal Resource Locator, (URL), and will retrieve the number of other sites that link to the given URL.

Dependent Variable

Price Premium: Price premium measures the amount that the total price of purchasing a product from a retailer exceeds the minimum price charged across all retailers for that specific good. Model results shown in Table 9A use total product price including shipping and handling cost; where the standard shipping option of 3-7 business days was chosen across all sites. Model results shown in Table 9B use product price premium, *not* including shipping and handling cost as the dependent variable. In some cases there was a price dispersion of up to $15.00 for the same standard shipping option, suggesting that some firms are attempting to hide the price premium they are charging in the shipping charge. This dispersion of standard shipping costs validates Brynjolfsson and Smith's 2000 finding that shipping price is negatively associated with repeat purchase behavior. Price information was gathered for each site by the research team at the same time that the qualitative rating assessments were gathered.

A Correlation matrix of variables, as well as means and variances is shown in Table 5. In addition, the magnitudes of price dispersion across product categories are shown in Table 6.

_____see Table 5 and Table 6_____

Data Analysis

In this section we describe the key data analysis procedures.

Model Specification

We analyze a linear specification of the model, allowing us to explore the additively separable and linear effects of e-business factors on price premium. Note that we control for retail price in model (2), to explore price premium relative to the good's retail price. The empirical models, representing two stages of research, are shown below:

$$\textbf{Personalization} = \alpha_1 + \alpha_2*(\textbf{User Interaction}) + \alpha_3*(\textbf{Product Information}) + \alpha_4*(\textbf{Ease of Use}) + \alpha_4*(\textbf{Brand}) + \alpha_5*(\textbf{Book Type/CD Type/PDA Type}) + \alpha_5*(\textbf{Number of Links}) + \varepsilon_1 \quad (1)$$

Price Premium = α_1 + α_2*(**Personalization**) - α_3*(**Product Information**) + α_4*(**User Interaction**) + α_5* (**Ease of Use**) + α_6*(**Shipping and Handling**) +α_7*(**Brand**) + α_8*(**Retail Price**) +α_9*(**Number of Links**) + ε_1 (2)

Books, CDs, and PDAs are significantly different across various product attributes and customer buying frequencies; therefore, we tested for structural differences across these product characteristics using the Chow test (Chow 1960). The results of the Chow test for CD ratings and book ratings produced an F-statistic of 9.46 ($p < 0.0005$); such a result asserts a structural difference in the linear model with the pooled CD and book data. The Chow test for CD ratings and PDA ratings produced an F-statistic of 12.40 ($p < 0.0005$); such a result asserts a structural difference in the linear model with the pooled CD and PDA data. The results of the Chow test for CD ratings and PDA ratings produced an F-statistic of 17.44 ($p < 0.0005$); such a result asserts a structural difference in the linear model with the pooled book and PDA data. Consequently, the data across three product categories were analyzed using separate models since the Chow test rejected the null hypothesis that data across product categories were structurally similar.

Estimation Procedures

The model parameters were estimated using ordinary least squares ("OLS"). Standard assumptions of these estimators were tested. In testing for heteroskedasticity, White's Test (White 1980) produced a p-value of 0.4022 for model (2), the price premium model, indicating heteroskedasticity is not an issue, as we fail to reject the null hypothesis. In the case of model (1), White's test produced a p-value of 0.019; therefore, we reject the null hypothesis. To resolve the issue of heteroskedasticity in model (1) we used White's corrected standard errors (White 1980). The resulting regression showed no change in the coefficients, only the residuals.

Multicollinearity refers to significant correlations among the independent variables. In our model, it may be argued that independent variables such as raw product information and personalization are correlated. Hence, it is important to test for any significant multicollinearity effects on the parameters in our models. The effect of multicollinearity in the above model was evaluated by computing the Variance Inflation Factor ("VIF") for each independent variable. The mean VIF for the two models in our analysis were 2.33 and 1.96 respectively, indicating no significant effect of multicollinearity on parameter estimates. An alternate approach is to check for the condition number; a large condition number indicates multicollinearity. Some declare significant multicollinearity if the condition number exceeds 30 (Weisberg, 1985). In our analysis, the highest condition number for the personalization model is 4.2221, while the highest condition number for the price premium model is 3.5593; this indicates no evidence of severe multicollinearity in our models. The resulting parameter estimates of the models are shown in Table 7 and 9A and 9B. The calculated values of F-statistics for the models exceeded the critical values at the 5% significance level; indicating that our models explain a significant portion of the variance in Personalization and Price Premium, respectively. The results for Model 2, the price premium model, are shown with two dependent variables: A) shipping cost included; and b) shipping cost not included.

We also computed the Cook's distance for each data point in our sample to check the influence of any outliers in our parameter estimates (Cook and Weisberg, 1982). The maximum Cook's distance in our sample for model (1) was .057, and for model (2) was .113. Such results indicate the absence of any single influential data point. The results for the personalization model and the price premium model are shown below; they are listed for each product category separately in Table 7 for model (1) (personalization) and in Table 9A and 9B for model (2) (price

premium). Since it is possible that a common omitted variable could have affected the error terms in both models, we estimated their parameters using Seemingly Unrelated Regression Estimation ("SURE"). The SURE estimates were not significantly different from OLS estimates, which indicate the absence of any significant correlation across the error terms of two models.

____See Table 7____

6) Results and Discussion

The parameter estimates of model (1) with the t-statistics are presented in Table 7. As expected, the positive sign of the significant **Product Information** coefficient across all three product categories means that increased product information is positively associated with increased personalization. Thus, controlling for more popular firms with potentially larger consumer databases, increased amounts of product information is still associated with increased personalization; (H1) is therefore supported. **User Interaction**, on the other hand, was not significant in the personalization model. Thus, while there is a positive association between user interaction and personalization, the result is not significant; H2 therefore, is not significantly supported.

In regards to the control variables in the personalization model, several interesting results are seen. First, the positive signs of the significant **Ease of Use** coefficients across all three product categories indicate that an easy to use website is associated with a greater amount of personalization provided by the website. This result makes sense, as personalization likely makes a website easier to use, that the association is expected. In addition, **Brand** is significant across the book retailers and the PDA retailers, but with differing signs. The positive sign of **Brand** across only book retailers indicates that a branded retailer is associated with more personalization than an unbranded retailer, although the magnitude of the brand coefficient is not

that large. Conversely, the negative sign of **Brand** across only PDA retailers indicates that a branded retailer is associated with less personalization than an unbranded retailer. This result is aligned with Peterson et al.'s product classification scheme, as PDAs are higher priced and less frequently purchased goods relative to books and CDs. Thus while consumers may repeatedly buy books in a given period of time, most consumers do not repeatedly buy PDAs; hence, the need for frequent interactions may not exist. Also, at the time of our study there were a limited number of PDA types; the relatively smaller number of substitutes in the PDA market means that there was less data from which to make purchase recommendations. In addition, PDAs are more complex products that may require user understanding of the various features as well as details in compatibility with various computer applications. Therefore, PDA shoppers most likely already know what product they want to buy before entering and thus have less need for personalized recommendations of best-fit products.

The positive sign of **Number of Links,** significant across all three product-types, indicates that an increase in the number of external web sites that link to a firm is associated with greater personalization. Number of links is a proxy-variable for popularity or market presence. Thus, a site's increased popularity or market reach is associated with increased personalization. Personalization technologies, such as collaborative filtering, are more effective with increased amounts of both product and consumer data. Thus, more popular sites with a greater market presence likely have more consumers in their database and are therefore able to offer more personalization.

A comparison of the results is shown below in Table 8.

____See Table 7 and Table 8____

The parameter estimates of the price premium model (2) with the t-statistics are presented in Tables 9A (price premium when shipping is included) and 9B (price premium when shipping is not included). The results across the two dependent variables are largely consistent. H3 is supported in the book and PDA industry, as **Personalization** is positive and significantly associated with price premium; thus a higher level of personalization is associated with a greater price premium in the book and PDA industries. As theorized in hypothesis three, increased personalization likely decreases within-site product information search costs for the consumer and therefore enhances the firm's ability to extract a surplus from the consumer through a price premium. In the CD industry, on the other hand, the **Personalization** coefficient is negative and significant, indicating that a higher amount of personalization is associated with a lower price premium for *CD* retailers; thus, H6 is supported. A negative association of personalization and price premium in the CD industry is likely due to the greater variance of customer expectation of music, and "experience goods" in general, as has been suggested by previous research (Gu and Hitt, 2001). As a consequence, it may be difficult to effectively personalize CD offerings and make recommendations that meet specific consumer needs. Hence, personalization may not be of much value to consumers in the CD industry; in fact, it can be argued that personalization in CDs may even increase consumer search cost as consumers are potentially shown offerings that are not of interest to them.

The significant negative sign of the **Product information** coefficient across the product categories indicates that an increase in the dimensions of available product information is associated with price premium decreases; thus, H4 is supported. This result is consistent with the idea that increased dimensionality of information increases product information search cost: consumers have more information to process in order to determine the best-fit product. In other

words, a firm that provides a greater number of dimensions of product information increase within-site product information search costs, and therefore makes it difficult for consumers to meet their informational needs. This result may also indicate that un-personalized product information may lead to information overload, and thereby affect the quality of consumer decisions. The increased consumer effort needed to determine the best-fit product is likely associated with a decrease in price premium, as firms are not able to extract the optimal surplus amount from the consumer. This result is consistent across both dependent variables, except the result is not significant in the CD industry when shipping cost is not included in the total price. Thus, it is possible in the CD industry, since prices are rather standard for CDs, that more of the price dispersion is hidden in the shipping cost.

One might suspect that a nonlinear relationship exists between price and information, such that both too much information and too little information increases product information search costs. We tested for a nonlinear relationship in our model by including a quadratic information term. However, the term was not significant, thus we omitted it from the model. We suspect that the quadratic term was not significant due to the constrained scope of our information measure being on a 0-6 scale.

The significant positive sign of the **User Interaction** coefficient for the book and CD industry when shipping cost is included, and in the book industry when shipping cost is not included, indicates that increased useful user interaction is associated with an increase in price premium; thus H5 is weakly supported. User interaction in this study is the ability to interact with stored pending transactions or previous order history. The positive association of user interaction to price is likely because stored pending transactions cause an increase in consumer cross-firm price information search cost. For example, when consumers access their stored

shopping cart they are not presented with alternative pricing options at other stores. Consequently, the increase in price search cost may decrease price competition between firms, allowing firms to charge price premiums due to increased market power. On the other hand, the effect of **User Interaction** on PDA price premiums is not significant in either model, and the effect is only significant in the CD category when shipping cost is included. In the case of the CD industry, we suspect that, as mentioned above, a portion of the price dispersion is hidden in the shipping costs, and thus the **User Interaction** effect is not strong enough to be significant when shipping costs are not included. In the PDA industry, on the other hand, we suspect that the lack of a significant result is due to the classification of PDAs as high-cost, low-frequency purchases; thus PDA buyers at the time of the study were largely one-time buyers rather than repeat buyers and did not therefore value features that increase interaction over multiple shopping visits.

The significant negative sign of the **Ease of Use** coefficient for the book industry indicates that firms with websites perceived as user-friendly are associated with a decrease in price premium. This finding likely speaks to the resources of larger firms who possess economies of scope and the knowledge base to implement a user-friendly website. Simultaneously, these larger firms possess the economies of scale to allow them to minimize price premiums and attract new consumers. The result is opposite in the CD industry, where the significant positive sign of the **Ease of Use** coefficient indicates that firms with websites perceived as easier to use are associated with an increase in price premium. This may be due to the fact that at the time of our study, dominant online retailers such as Amazon.com had just entered the CD industry; there were no dominant CD online retailers to create a significant price war comparable to that of the book industry. Such a lack of dominant CD online retailers is also

reflected in the absence of **Brand** effect on price premium in the case of CDs. Another possible explanation for the significant positive sign of the **Ease of Use** coefficient is that most technically savvy Internet shoppers download digital music; thus, online CD consumers are largely less technologically sophisticated. As a result, ease of use becomes a competitive advantage for CD retailers online, enabling them to charge higher price premiums. In addition, online music ranges across a wide span; in fact, often consumers experience only one song of an artist, and thus do not know the CD's title for which they are searching. Hence, ease of use becomes imperative for music shoppers searching for specific music ensembles.

The significant negative sign of the **Shipping and Handling** coefficient for the book and CD industry indicates that firms with websites rated with higher customer satisfaction regarding shipping and handling charge lower price premiums. Most often, consumers are satisfied with shipping and handling charges if they do not exist (see Brynjolfsson and Smith 2000). In other words, consumers would rate a firm higher for shipping and handling satisfaction if the firm charged lower prices. Shipping and handling charges are computed as part of the price premium in Table 9A but not in Table 9B; thus, firms with higher satisfaction in shipping and handling likely charge less overall in book and CD industry, as the results are consistent for these two industries whether or not shipping price is included in price premium. However, in the PDA industry, when shipping cost is not included, firms with greater shipping and handling satisfaction are associated with a greater price premium. Thus, firms with higher shipping and handling satisfaction charge a higher product price. The PDA industry had the large price dispersion for shipping and handling, at $15.00. Thus, in the case of the PDA industry, higher shipping and handling satisfaction is definitely associated with lower shipping prices, as shipping prices can make such a significant difference. It is worth noting that the shipping and handling

coefficient is based on consumer surveys completed *after* receipt of the product. Thus, the data will only include the surveys of those who chose to complete them after receiving the product and may therefore include a time delay between the experience of receiving the product, and evaluating the shipping and handling.

The **Brand** variable is not significantly associated with price premium in the case of CD and PDA retailers in the price premium models; hence, it is difficult to directly assess whether there is indeed a distinct difference between branded and unbranded retailers in regards to price premium charged for PDAs and CDs. The significant negative sign of the **Brand** coefficient for the book industry indicates that firms with strong brand equity are associated with price premium decreases when shipping cost is included in the dependent variable. When shipping cost is not included, **Brand** is not significant; this suggests that branded retailers offer shipping promotions, or lower shipping costs, perhaps as an attempt to draw consumers to their site. It is also worth noting that during the project's data collection phase, the online retail market was in the post dot-com boom era, compelling large firms to decrease prices to lure consumers; therefore deterring smaller firms from entering the market. In addition, larger firms with the budget to create brand equity are likely benefiting from economies of scale, enabling them to charge lower prices. Considering this explanation, why, then, did such a result occur only in the book industry? One answer is that the CD and PDA markets appear to have more niche players and are thus less susceptible to these effects. Another explanation, supported by evidence, is that the immense market presence of the incumbents, Amazon.com (AMZN) and BarnesandNobles.com (B&N) in the online book industry has compelled other online retailers to target books and attempt to steal part of the incumbents' market share. For example, in June 2002, Buy.com announced that it would attempt to increase market share and lure consumers away from Amazon.com by offering

consumers an additional 10% off any book price listed on Amazon.com. In addition to the extra discount, Buy.com offered free shipping on all books, with no minimum order (Weiss 2002). Thus, shipping costs appear to be one of the first items to be reduced in an effort to attract consumers.

In summary, the findings in this field study provide evidence that online service factors such as **personalization**, **product information**, and **user interaction**, are associated with price dispersion online. Our results stress the need for online retailers to better streamline product information into useful customer information. Our study shows that an increase in the dimensionality of product information is actually associated with a decrease in price premium. It is well known that the scarcest resource online is consumer attention; thus, firms must be careful not to provide the "wrong" type of information. Our results suggest that when a firm presents too much raw product information, price premium suffers. Increased personalization, on the other hand, is associated with an *increase* in price premium. Personalization technologies use product information and attempt to present consumers with the "correct" type of information to determine their best-fit product. Hence, through appropriate personalization technologies, firms should strategically convert large quantities of product information into high quality information that meet a customer's contextual buying needs.

The issue of economies of scale in the post-dot-com-boom is one of notable importance. Price premiums from 1998 to 2001 have significantly decreased. The average price difference of the lowest price offered and the tenth lowest price offered in Brynjolfsson and Smith's 1998 data collection was $10.77 or 32.3%. The average price difference between the lowest price offered and the tenth lowest price offered for our 2001 data collection was $5.93, or 14.83%. The means that the two samples are significantly different at the $p < 0.00001$ level. An explanation for the

significant decrease in price premiums is that large companies may be utilizing economies of scale to force smaller players out of the market so they can subsequently increase price premiums closer to a monopolistic level.

Product Bundle Price Dispersion

While price premium in and of itself is not a sign of financial success, it can be a sign of consumer who are willing to pay more due to a greater service provided. The question then becomes across consumers, how substantial are these premiums at the point in time when we collected the data for this study. One may argue that prices may be higher for one product in a given store compared to another store, but lower for another product. Thus, since many shoppers purchase multiple items online (Farag and Krishnan, 2002), the next question becomes is there a systematic occurrence of a price premium in the electronic marketplace across sellers? If there is, indeed, a systematic difference, then the results of our previous analysis should stay consistent when product bundle price premiums are used as the dependent variable.

Thus, we constructed multiple random bundles, each comprising of 10 books and compared prices across these bundles. In doing this, we augmented our data set with price data of books not used in our initial analysis, but belonging to the same companies of our initial data set. The results of this analysis are shown in Table 11. Hypotheses 3 and 5 are supported in the product bundle regression, thus the positive significant association of **personalization**, as well as **user interaction**, with price premium exist even when product bundles are used to determine price premium. This result suggests that there is, indeed, a systematic price premium charged by some companies.

We fully acknowledge that even systematic price premium is not necessarily a sign of financial success. Thus, we gathered additional field data across 10 industries and assessed the correlation between personalization and conversion rate. We controlled for market presence using the number of Links metric. We found that personalization is positive and significantly associated with higher conversion rate of an online retailer, where conversion rate is defined as the number of consumers that purchased from a given site, divided by the number of consumers that visited the given site. This association is significant at the $p<0.003$ level. Thus, this result is further evidence that personalization is an important factor for the success of online firms.

Limitations

Our analysis is limited by the *recency* of our attempt to create metrics for online service factors such as personalization, product information, and user interaction. Future research should attempt to validate these metrics and further define objective metrics for them and other online service factors. Another limitation stems from our measurement of personalization, information, and user interaction as a *quantity* metrics, rather than a quality metric. In future studies, it would be enlightening to attempt to develop a metric for the *quality* of personalization, and test its association with price premium. In addition, a *quality* metric of user interaction would be of value, as some consumers may be put off from a web site with certain types of user interaction. In addition, it is possible that along with the ability to store pending transactions and view previous orders, other factors, such as access to live chat with customer service representatives, may also affect online consumer interaction. However, in our sample, we do not have reliable data on alternative measures and therefore leave such explorations for future research.

In addition, there are potentially variables that we do not include in the model, such as sales volume, that could contribute to price premium. In addition, we do not have a good metric for "customer information" that a firm may have. These variables were not included because we data for them were not readily available.

Conclusion and Future Research

To our knowledge, our analysis is the first to empirically study the association of online retailer IT features with price dispersion. In attempting to lay a foundation for quantifying the IT features an online retailer may implement, and their association with price, this research provides a significant contribution to the literature. Future research is required to define objective measures for other potential dimensions of e-business success that may affect price and market power.

Our field study of online retailer IT features provides several insights for e-business managers and other practitioners. First, our results provide evidence that there is a clear distinction between presenting large quantities of information to consumers and presenting personalized, consumer driven information. Given the existing literature on online information overload (ex: Adamic and Huberman 1999) and information asymmetry issues of price information (Salop and Stiglitz 1977), understanding the relationship between information and price online is an important issue. Our empirical evidence shows that increased product information is associated with a decrease in price premium. Such evidence suggests that firms should be interested in presenting users with relevant, or "proper" information, through personalization, rather than simply presenting the most information. Second, our analysis suggests that through user interaction features (such as saved shopping carts and previous order histories); consumers cross-firm price information search costs increase, and firms are able to

therefore charge a systematic price premium for a homogenous good. In other words, by giving consumers the convenience of storing items of interest, firms may decrease consumer comparison-shopping and charge a higher price.

Our study is the first to empirically show that personalization is associated with higher prices. In addition, this study is the first to empirically contrast the effects of personalization across three product categories of varying degree of customer heterogeneity of preferences. In industries where consumer preferences are relatively uniform, such as the book and PDA industry, firms should utilize personalization to present better quality and proper consumer information that will allow firms to extract consumer surplus through potentially lower information search costs. Such scenarios of increased consumer information quality may in-turn provide firms increased market power, as evidenced by the systematic price dispersion across a bundle of books. Our study suggests that firms must understand the product category they are selling, and the relative variance of consumer preferences within the category. Online retailers can then utilize careful implementation of personalization, information, and user interaction to turn their ventures into entities of significant market power, where profitability exists around the corner, despite the perceived extinction of the dot-com boom.

References

1. Adamic, L., Huberman, B. A. (1999). "The Nature of Markets in the World Wide Web." *Computing in Economics and Finance*, Meetings of the Society for Computational Economics.
2. Alba, J., Lynch, B., Janiszewski, C., Lutz, R., Sawyer, A., and Wood, S. (1997). "Interactive home shopping: Consumer, retailer, and manufacturer incentives to participate in electronic markets." *Journal of Marketing* **61**(3): 227-247.
3. Bakos, J. Yannis (1997). "Reducing Buyer Search Costs: Implications for Electronic Marketplaces." *Management Science* **43**(12).

4. Baron, R. M. and Kenny, D. A. (1986). "The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic and Statistical Considerations". *Journal of Personality and Social Psychology*, 51(6): 1173-1182.
5. Barua, A. and C.H. Kriebel; T. Mukhopadhyay (1995). "Information technologies and business value: An analytical and empirical investigation." *Information Systems Research* **6**(1): 3-23.
6. Berscheid, Ellen and Walster, Elaine (1978). *Interpersonal Attraction*. Reading, MA, Addison-Wesley.
7. Blattberg, Robert C., and Neslin, Scott A. (1990). *Sales Promotion: Concepts, Methods, and Strategies*. Englewood Cliffs, NJ, Prentice Hall.
8. Bradlow, Eric T. and Schmittlein, David C. (2000). "The Little Engines That Could: Modeling the Performance of World Wide Wed Search Engines." *Marketing Science* **19**(1): 43-62.
9. Brynjolfsson, Erik; Smith, Michael (2000). "Frictionless Commerce? A Comparison of Internet and Conventional Retailers." *Management Science* **46**(4): 563-585.
10. Chow, Gregory C. (1960). "Tests of Equality between Sets of Coefficients in Two Linear Regressions." *Econometrica* **28**(3): 591-605.
11. Clay, Karen, Krishnan, Ramayya and Wolff, Eric (2001). "Prices and Price Dispersion on the Web: Evidence from the Online Book Industry" *Journal of Industrial Economics* **49**(4): 521-535.
12. Clemons, Eric, Hann, Il-Horn, Hitt, Lorin (2002). "Price Dispersion and Differentiation in Online Travel: An Empirical Examination." *Management Science* **48**(4): 534-549.
13. Cook, R. D. and S. Weisberg, *Residuals and Influence in Regression*, Chapman & Hall: London, 1982.
14. Curtis, Rebecca S. and Miller, Kim (1986). "Believing Another Likes or Dislikes You: Behavior Making the Beliefs Come True." *Journal of Personality and Social Psychology* **50**(August): 284-190.
15. Daft, R. and Lengel, R. (1986). "Organizational Information Requirements, Media Richness, and Structural Design." *Management Science* **32**: 554-571.
16. Dahlby, B., and West, D. S. (1986). "Price Dispersion in the Automobile Insurance Market." *Journal of Political Economy* **94**(2): 418-438.
17. Dana, James (2001) "Competition in Price and Availability when Availability is Unobservable", *the RAND Journal of Economics,* Vol. 32, No. 3, pp. 497-513.
18. Diamond, P.A. (1985). Search Theory. *New Pelgrave*.
19. Doll, William J. and Torkzadeh, Gholamreza (1988). "The Measurement of End-User Computing Satisfaction." *MIS Quarterly* **12**(2): 258-274.
20. Drachman, David, de Carufel, Andre, and Insko, Chester A. (1978). "The Extra Credit Effect in Interpersonal Attraction." *Journal of Experimental Social Psychology* **14**(September): 458-467.
21. Erevelles, S., Rolland, E., Srinivasan, S. (2001). "Are Prices Really Lower on the Internet? An Analysis of the Vitamin Industry." *Working Paper, University of California, Riverside*.
22. Fader, Peter S., and Bruce G. Hardie (1999) "Forecasting Repeat Sales at CDNOW: A Case Study." *Working Paper, The Wharton School of the University of Pennsylvania.*
23. Farag, Neveen I. and Van Alstyne, Marshall W. "Online Personalization: Bringing the Value to Consumers or the Consumers to the Value?" *UMBS Working Paper*, 2001.

24. Fishman, Arthur (1992). "Search Technology, Staggered Price-Setting, and Price Dispersion." *The American Economic Review* **82**(March): 287-298.
25. Futrell, Charles (1988). *Fundamentals of Selling*. Homewood, IL, Irwin.
26. Gazzale, Robert S. and Jeffrey MacKie-Mason (2001). "Endogenous Differentiation of Information Goods under Uncertainty." *Proceedings of Electronic Commerce (EC-01).* ACM Press, Miami, FL.
27. Giuletti, Monica (1999). "Price Discrimination in Grocery Trade: Evidence From Italy." *Applied Economics* **31**: 319-329.
28. Gu, Bin and Lorin M. Hitt (2001) "Transactions Costs and Market Efficiency," *22nd Annual International Conference on Information Systems*, New Orleans, LA.
29. Hall, Robert E. (1988). "The relation between Price and Marginal cost in U.S. Industry." *The Journal of Political Economy* **96**(5): 921-947.
30. Harrington, Joseph E. (2001). "Comment on "Reading Buyer Search Costs: Implications for Electronic Marketplaces"." *Management Science* **47**(12): 1727-1732.
31. Hoffman, Donna L. and Thomas P. Novak (1995) "Marketing in Hypermedia Computer-Mediated Environments: Conceptual Foundations," Project 2000 Working Paper No. 1. Owen Graduate School of Management, Vanderbilt University.
32. Hoffman, D., Novak, T., and Chatterjee, P. (1995). "Commercial Scenarios for the Web: Opportunities and Challenges." *Journal of Computer Mediated Communication* **1**(3): <www.ascusc.org/jcmc/vol1/issue3/hoffman.html.
33. Ives, B., Olson, M., and Baroudi, S. (1983). "The Measurement of User Information Satisfaction." *Communications of the ACM* **26**(10): 785-793.
34. Iyengar, Sheena and Lepper, Mark R. (2000), -- "When Choice Is Demotivating: Can One Desire Too Much Of A Good Thing?" Journal Of Personality And Social Psychology, 78, 995-1006.
35. Jacobs, Larry, Berscheid, Ellen, and Walster, Elaine (1971). "Self-Esteem and Attraction." *Journal of Personality and Social Psychology* **17**(January): 84-91.
36. Johnson, Carrie A., Allen, Lisa, and Scheirer, Lisa (2001). "Digital Music Will Squeeze Physical CD Sales." *Forrester TechStrategy Brief.*
37. Kennedy, Peter (1998). *A Guide to Econometrics*. Cambridge, MA, The MIT Press.
38. Kerlinger, F.N. (1978). *Foundations of Behavioral Research*. New York, McGraw-Hill.
39. Kim, W. C., R. Mauborne (1997). "Value innovation: The strategic logic of high growth." *Harvard Business Review*. **75**: 103–112.
40. Kroll, Karen M. (2001). "You're Gonna Love This." *Catalog Age* **April**: S1-S4.
41. Levy, Michael and Weitz, Barton (1992). *Retailing Management*. Homewood, IL, Irwin.
42. Lynch, John G. Jr., Ariely, Dan (2000). "Wine Online: Search Costs Affect Competition on Price, Quality, and Distribution" *Marketing Science* 19(Winter: 1): 0083-0103.
43. Marks, Ronald B. (1991). *Personal Selling: An Interactive Approach*. Needham Heights, MA, Simon & Schuster.
44. Mascarenhas, B. (1992). "First Mover Effects in Multiple Dynamic Markets." *Strategic Management Journal* **13**(March): 237-243.
45. Nelson, P. (1970). "Information and Consumer Behaviour." *Journal of Political Economy* **78**: 311-329.
46. Nunnally, J. (1967). *Psychometric Theory*. New York, McGraw-Hill.
47. Nunes, Paul F., and Kambil, Ajit (2001). "Personalization? No Thanks." *Harvard Business Review* **79**(4): 32-34.

48. Palmer, J. (1997). "Retailing on the WWW: The use of electronic product catalogs." *International Journal of Electronic Markets* 7(3): 6-9.
49. Peterson, Robert A., Balasubramanian, Sridar, Bronnenberg, Bart (1997). "Exploring the Implications of the Internet for Consumer Marketing." *Journal of the Academy of Marketing Science* **24**(4): 329-346.
50. Pratt, John W.; Wise, David A. and Zeckhauser, Richard (1979). "Price Differences in Almost Competitive Markets." *The Quarterly Journal of Economics* **93**(2): 189-211.
51. Ratchford, Brian T.; Agrawal, Jagdish; Grimm, Pamela; and Srinivasan, Narasimhan (1996). "Toward Understanding the Measurement of Market Efficiency." *Journal of Public Policy and Marketing* **15**: 167-184.
52. Regan, Judith W. (1976). "Liking for Evaluatiors: Consistency and Self-Esteem Theories." *Journal of Experimental Social Psychology* **12**(March): 159-169.
53. Salop, S., and J.E. Stiglitz (1977). "Bargains and Ripoffs: A Model of Monopolistically Competitive Price Dispersion." *The Review of Economic Studies* **44**: 493-510.
54. Shaffer, G. and Zhang, J. (1995). "Competitive Coupon Targeting." *Marketing Science* **14**(4): 395-416.
55. Shardanand, U., and P. Maes (1995). "Social Information Filtering: Algorithms for Automating 'Word of Mouth' ". *CHI' 95 Mosaic of Creativity.*
56. Smith, Michael D.; Bailey, Joseph P.; Brynjolfsson, Erik (2000). Understanding Digital Markets: Review and Assessment. *Understanding the Digital Economy.* Erik Brynjolfsson and Brian Kahin. Cambridge, MA, MIT Press.
57. Sorenson, Alan T. (2000). "Equilibrium Price Dispersion in Retail Market for Prescription Drugs." *Journal of Political Economy* **108**(4): 833-850.
58. Stigler, George (1961). "The Economics of Information." *Journal of Political Economy* **69**(3): 213-225.
59. Stiglitz, J.E. (1989). Imperfect Information in the Product Market. *Handbook of Industrial Organization.* R. Schmalansee and R. D. Willig (eds). North Holland, New York: 769-847.
60. Thisse, J.F. and Vives, X. (1988). "On the Strategic Choice of Spatial Price Policy." *American Economic Review* **78**: 122-137.
61. Weill, Peter, and Michael R Vitale (2001). Making Clicks Become Bricks. *Place to Space: Migrating to eBusiness Models*, HBS Press.
62. Weisberg, S. (1985) *Applied Linear Regression.* New York: John Wiley and Sons.
63. Weiss, Todd (2002). Buy.com challenges Amazon.com on online book prices. *Computerworld.com*: http://www.computerworld.com/industrytopics/retail/story/0,10801,72241,00.html.
64. Witsman, Karl (1987). "No More What's His Name Again." *American Salesman* **32**(February): 24-27.
65. White, H.(1980) "Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica* **48**(5): 817-838.

Figure 1: Expected Results

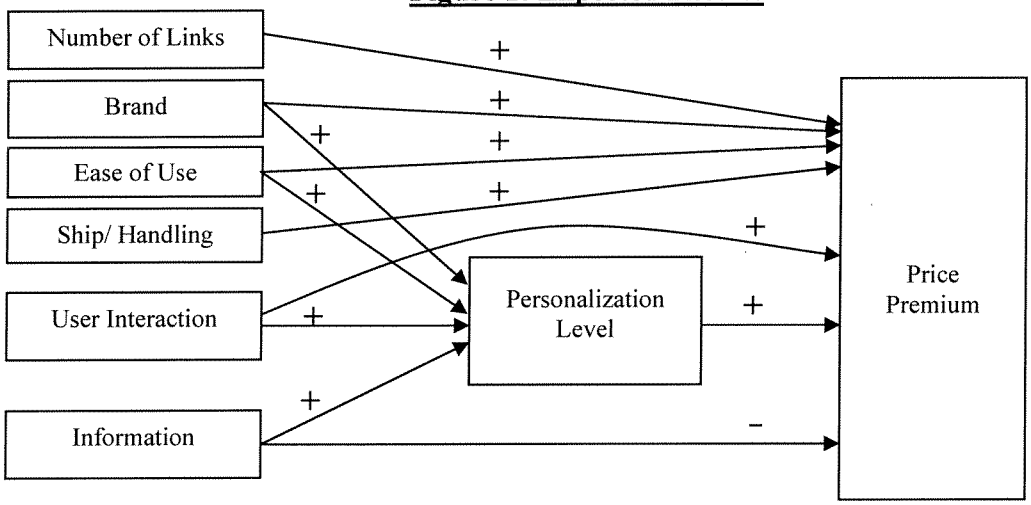

Table 1: Summary of Variables

Construct	Description of Definition	Code or Measures
Personalization		Average across score of 5 IT feature counts 1. Number of Time User was referred to by name upon logging into the site 2. The number of new features presented to user after the user logs in (wish list, send to a friend, etc.) 3. The total number of links that changed aligned with user revealed preferences upon login 4. The total number of customers reviews presented to the consumer of each product 5. The total number of customer ratings of reviews that were presented to the customer
User Interaction	The amount of transaction-based interaction afforded to a consumer	1. Can a customer store products in a shopping cart? (0/1) 2. Can a customer store and access previous order information (credit card used, shipping address, products purchased)? (0/1)
Product information	The amount of product information available	1. A list of 4 or more product attributes 2. Table of contents 3. Book excerpt 4. Reviewer Information 5. Site-provided product review 6. Full product pricing information on product page (rather than partial price information until the product is put in the shopping cart)

Table 2: Cronbach's Alpha Reliability

Variable	Standardized Reliability
Personalization	.90
User Interaction	.76

Information	.88

Table 3: Factor Analysis

	Personalization	User Interaction	Information
By Name	**1.02452**	0.02405	-0.06383
Func Change	**1.02812**	0.02029	-0.05539
Pers Links	**0.61043**	-0.35916	0.25557
Customer Reviews	**0.60240**	0.02019	0.45666
Rat Cust Reviews	**0.65312**	0.02367	0.28640
ShopCart	-0.03479	**0.83561**	0.08348
OrderHistory	0.12038	**0.50210**	0.08572
Number of Info	0.03771	0.04712	**0.93198**
Table of Contents or List of Tracks	0.06563	-0.01066	**0.85184**
Excerpt Of Book or CD	-0.13764	-0.03355	**0.74410**
Reviewer Info	0.09102	0.02713	**0.73108**
Site Review	0.20382	0.25353	**0.60623**
Product Price	0.02269	-0.10526	**0.45743**

Table 4: Coding of Brand Variable

1	Online only book/CD/PDA specialist
2	Online only book/CD/PDA generalist
3	Online only generalist
4	Offline Branded Retailer

Table 5 Summary Statistics and Correlation Matrix

	Mean	Std. Dev.							
Personalization	0.18	0.20	1.00						
User Interaction	0.82	0.28	0.26	1.00					
Information	0.36	0.20	0.68	0.36	1.00				
Number of Links	18017.93	32168.03	0.58	0.40	0.79	1.00			
Brand	0.30	0.46	0.60	0.49	0.60	0.71	1.00		
Ease of Use	8.92	0.34	0.26	-0.20	0.24	0.14	0.02	1.00	
Shipping and Handling Satisfaction	8.05	0.73	-0.14	-0.23	-0.07	-0.18	-0.37	0.42	1.00

Table 6: The Magnitudes of Price Dispersion across Product Categories

	Books		CDs		PDAs	
	Product Only	Product & Shipping	Product Only	Product & Shipping	Product Only	Product & Shipping
Mean	3.17	4.93	2.34	2.52	32.22	29.80
Std.	4.05	3.98	1.37	1.56	22.69	21.65

Dev.						
Min	0.00	0.00	0.00	0.00	0.00	0.00
Max	18.00	19.06	5.38	6.06	96.67	95.81

Table 7 Personalization Model (1) Parameter Estimates
Personalization across Book/CD/PDA Retailers

Variable	Parameter	Books	CD	PDA
Intercept	α_1 (t-statistics)	-1.34639** (-2.93)	-1.3835** (-3.26)	-1.3123** (-4.01)
Information	α_2 **(t-statistics)**	**0.49696** (4.17)**	**0.33672** (4.78)**	**0.4811** (6.82)**
User Interaction	α_3 (t-statistics)	0.08508 (1.52)	-0.00956 (-0.63)	0.02368 (1.40)
Ease of Use	α_4 (t-statistics)	**0.12944** (2.72)**	**0.17703** (3.45)**	**0.15166** (4.24)**
Brand	α_5 (t-statistics)	**0.03138* (1.91)**	-0.00449 (-0.26)	**-0.02736 ** (-3.84)**
Number of Links	α_6 (t-statistics)	**1.52e-06** (2.00)**	**1.69e-06** (4.33)**	**0.0000546** (5.59)**
Book /CD/PDA Type	α_7 (t-statistics)	0.00676 (0.87)	-0.00576 (0.50)	0.0002 (0.03)
Adj. R^2		0.4724	0.5187	0.4959
F-stat		25.02 ($p < 0.0001$)	24.89 ($p < 0.0001$)	17.89 ($p < 0.0000$)

** 5% level of significance *10% level of significance

Table 8: Personalization Model Comparisons

	User Interaction	Information	Ease of Site	Brand	Book/CD/PDA type	Links in
Book	ns	+ **	+ **	+ *	Ns	+ **
CD	ns	+ **	+ **	Ns	Ns	+ **
PDA	ns	+ **	+ **	-**	Ns	+ **

ns- Not Significant, ** 5% level of significance, *10% level of significance

Table 9A Price Premium Model (2) Parameter Estimates
Total Price Premium (Including Shipping) across Book/CD/PDA Retailers

Variable Name	Parameter	Books	CDs	PDAs
Intercept	α_1	24.610**	-14.6889**	24.0723
	(t-statistics)	(3.18)	(-2.27)	(1.28)
Personalization	α_2	3.7902**	-3.2877**	10.7517**
	(t-statistics)	(2.04)	(-3.44)	(2.07)
Information	α_3	-7.5335**	-2.2330**	-9.6695**
	(t-statistics)	(-3.06)	(-2.63)	(-2.16)
User Interaction	α_4	4.8828**	0.4312**	-0.3829
	(t-statistics)	(4.50)	(2.54)	(-0.41)
Ease of Use	α_5	-1.7905*	2.9711**	-3.4462
	(t-statistics)	(-1.90)	(4.95)	(-1.52)
Shipping and Handling	α_6	-0.9124**	-1.4776**	0.6178
	(t-statistics)	(-2.08)	(-2.83)	(0.76)
Brand	α_7	-2.2830**	-0.00262	0.04839
	(t-statistics)	(-2.45)	(-0.01)	(0.12)
Retail Price	α_8	0.08329**	0.2460**	0.0255**
	(t-statistics)	(4.37)	(2.21)	(6.57)
Number of Links	α_9	0.000023	0.0000114**	-0.0000773
	(t-statistics	(1.60)	(2.12)	(-0.12)
R^2		0.2993	0.3740	0.3125
F-stat		9.54	10.86	6.80
		($p < 0.000$)	($p < 0.000$)	($p < 0.000$)

** 5% level of significance * 10% level of significance

Table 9B Price Premium Model (2) Parameter Estimates
Product Price Premium (without shipping costs) across Book/CD/PDA Retailers

Variable Name	Parameter	Books	CDs	PDAs
Intercept	α_1	15.06*	-23.66**	170.05*
	(t-statistics)	(1.83)	(-4.10)	(1.87)
Personalization	α_2	3.23*	-2.54**	44.27*
	(t-statistics)	(1.94)	(-2.97)	(1.96)
Information	α_3	-7.27**	-1.05	-35.64*
	(t-statistics)	(-2.77)	(-1.41)	(-1.74)
User Interaction	α_4	3.44**	0.23	0.16
	(t-statistics)	(2.98)	(1.60)	(0.04)
Ease of Use	α_5	-2.41**	3.01**	-24.12**
	(t-statistics)	(-2.44)	(6.13)	(-2.19)
Shipping and Handling	α_6	-0.9124**	-0.78*	9.55**
	(t-statistics)	(-2.08)	(-1.71)	(2.43)
Brand	α_7	0.8145	-0.0497	-0.46
	(t-statistics)	(1.75)	(-0.29)	(-0.24)
Retail Price	α_8	0.1010**	0.42**	0.02
	(t-statistics)	(4.98)	(4.20)	(0.92)
Number of Links	α_9	0.000023	3.91e-06	0.0016
	(t-statistics	(1.60)	(0.81)	(0.52)
Adj. R^2		0.2303	0.3397	0.1008
F-stat		6.98	9.49	2.43
		($p < 0.000$)	($p < 0.000$)	($p < 0.01$)

** 5% level of significance * 10% level of significance

Table 10 Price Premium Model (2) Parameter Estimates For Product Bundles

	Parameter	
Intercept	α_1 (t-statistics)	1311.1** (2.44)
Personalization	α_2 (t-statistics)	138.780** (3.32)
Information	α_3 (t-statistics)	18.28 0.55
User Interaction	α_4 (t-statistics)	373.73** (4.76)
Ease of Use	α_5 (t-statistics)	158.93** (2.99)
Shipping and Handling	α_6 (t-statistics)	-16.5787* (-1.98)
Brand	α_7 (t-statistics)	49.83** (3.14)
Number of Links	α_9 (t-statistics	-0.00023 -0.69
R^2		0.2289
F-stat		7.83 ($p < 0.000$)

** 5% level of significance * 6% level of significance