Girish Mallapragada, Sandeep R. Chandukala, & Qing Liu

Exploring the Effects of "What" (Product) and "Where" (Website) Characteristics on Online Shopping Behavior

Understanding factors that influence online shopping and managing consumer relationships is not a trivial task for firms, considering the many pertinent factors that influence behavior, including the product being shopped (i.e., the "what") and the context of the website itself (i.e., the "where"). This study investigates the impact of these characteristics on an online transaction's basket value, after incorporating the role of other aspects of the browsing process including page views and visit duration. The authors estimate a multivariate mixed-effects Type II Tobit model with a system of equations to explain variation in shopping basket value, using data involving 773,262 browsing sessions resulting in 9,664 transactions across 43 product categories from 385 unique websites. The results support the assertions that contextual factors are associated with online browsing. For example, a website's scope in terms of product variety is associated positively with visit durations and basket values but negatively with page views. Furthermore, a website's communication functionality is positively associated with basket value for hedonic products. Insights suggest managerial implications involving product and website strategies for online retailers.

Keywords: online retailing, multivariate mixed-effects models, product heterogeneity, basket value, website functionality

Online Supplement: http://dx.doi.org/10.1509/jm.15.0138

Online shopping offers a dominant alternative to traditional retail shopping and thus has garnered increasing interest from both practitioners and academics. Because online consumers' browsing history can be observed and recorded in detail, firms routinely seek to leverage these data to improve customers' experiences on their website. For example, Amazon.com offers customized recommendations based on users' browsing and purchase histories, and Apple's iTunes service provides recommendations according to previous media purchases. Managing customer experiences in a retail setting is highly contextual and important (Grewal, Levy, and Kumar 2009), but a nuanced understanding of the nature of online shopping outcomes and the factors that influence them remains an emerging area of inquiry (Kumar et al. 2013; Narayanan and Kalyanam 2015). With this study, we investigate how key outcomes of an online purchase experience—purchase decision and amount of money spent (hereinafter referred to as "basket value") might depend on the browsing characteristics (number of pages viewed and time spent browsing) and the characteristics of both the product category being shopped (i.e., the what) and the online retailer's website (i.e., the where), after controlling for consumer heterogeneity.

A typical online purchase experience includes multiple web page visits, through which the consumer processes the gathered information, before eventually making a purchase. A consumer's visit to a website is described using a session, conceived as starting when a consumer visits a specific website and ending when either the consumer leaves the website or closes the browser tab or window. The following outcomes of a shopping experience might occur within any given browsing session: page views, visit duration, decision to purchase, and, conditional on the decision, basket value of the purchase. To understand the drivers of these outcomes, retailers need to observe the behavior of shoppers in various contexts (e.g., Kushwaha and Shankar 2013; Raghu et al. 2001). Online retailers are relatively limited in their ability to analyze consumers' browsing and purchase behaviors comprehensively, because they observe browsing and transactions only on their own websites. Furthermore, manufacturers need to understand how a retailer's website characteristics or product mix influence their own products while on display,

Girish Mallapragada is 3M Faculty Fellow and Assistant Professor of Marketing, Kelley School of Business, Indiana University (e-mail: gmallapr@indiana.edu). Sandeep R. Chandukala is Associate Professor of Marketing, Lee Kong Chian School of Business, Singapore Management University (e-mail: sandeepc@smu.edu.sg). Qing Liu is Associate Professor of Marketing, Wisconsin School of Business, University of Wisconsin (e-mail: qliu@bus.wisc.edu). The authors gratefully acknowledge feedback from Gary L. Lilien and Rajdeep Grewal on a previous version of this article. Rajkumar Venkatesan served as area editor for this article.

ISSN: 0022-2429 (print)
1547-7185 (electronic)

Journal of Marketing
Vol. 80 (March 2016), 21–38
DOI: 10.1509/jm.15.0138

as well as how these effects might vary across retailer websites. Accordingly, we posit that the online shopping outcomes depend on heterogeneity in what is being shopped (product characteristics) and where it is being shopped (website characteristics), as well as their potential interactions, after controlling for consumer heterogeneity (Yadav and Pavlou 2014).

Insights from extant academic research regarding the interrelated nature of online shopping outcomes and the role of various contextual factors are insufficient, however, because of the practical constraints arising from issues of scale and scope. First, most research investigates browsing and purchasing online behaviors separately, ignoring their co-occurrence within a single browsing session (e.g., Johnson et al. 2004). Second, extant studies typically explore online behavior by focusing on one or a limited number of retailers (e.g., Sridhar and Srinivasan 2012), a few aspects of the browsing process (e.g., Danaher, Mullarkey, and Essegaier 2006), or a limited number of product categories (e.g., Montgomery et al. 2004). Such constrained approaches can produce rich insights about a specific issue within the online shopping experience or for a specific category, but they are limited in generalizability across product categories or retailers. Some authors attempt to be more general (e.g., Zhu and Zhang 2010) but remain restricted by the size of their analysis sample or the limited scope of the problem.

To extend the literature in the domain of online shopping experiences, we propose an integrative approach to account for a staged purchase process with multiple outcomes estimated by a multivariate mixed-effects model (e.g., Lindstrom and Bates 1990). Specifically, we model the notion that online shopping involves browsing that may result in a purchase decision, conditional on which a basket of certain value is purchased. In addition, we jointly investigate the impact of product and website characteristics on the outcomes after controlling for consumer heterogeneity.

We assemble a comprehensive data set of hundreds of thousands of online browsing sessions across hundreds of online retailers and numerous product categories, which helps alleviate limitations due to restricted sample sizes and problem scopes. Furthermore, we supplement these transaction data with additional data about the characteristics of product categories and the functionality of online retailer websites using multiple methods. Finally, we specify the trade-offs between browsing and transaction outcomes, as well as the synergistic effects of product and website characteristics. Tests of several alternate specifications and assumptions of our proposed model demonstrate the robustness of our findings.

The results offer notable insights for both online retailers and manufacturers selling through those retailers. For example, the number of pages viewed is positively associated with purchase decision and basket value, whereas visit duration is positively related to purchase decision but not to basket value. In terms of product characteristics, online retailers with a broad variety of product categories tend to benefit more than retailers with a narrow variety when the products have more hedonic or utilitarian traits. We also find that communication functionality (e.g., chat room, messaging) of the online retailer's website is positively associated with basket value when consumers shop for products that are more hedonic, whereas the navigational functionality (e.g., feedback, site maps) is beneficial when products are more utilitarian. Finally, the factors that are positively related to page views and visit durations do not necessarily have the same association with basket value.

In the next section, we begin by describing the consumer browsing and purchase process. Next, we motivate our empirical approach by elaborating on the contextual factors that drive the key outcomes, focusing on the nature of heterogeneity. After we describe our data, we develop our multivariate mixed-effects model and present our results. Finally, we discuss our key findings, focusing on managerial insights derived from our proposed model, as well as some study limitations and opportunities for further research.

Contextual Background

Consumer Browsing and Purchase Process

Research on online shopping behavior offers various measures of consumer engagement. Two important, widely used measures are page views (Danaher 2007; Huang, Lurie, and Mitra 2009) and duration of the visit (Danaher et al. 2006; Montgomery et al. 2004). Although page views and visit duration are positively associated with consumer choice (Lin et al. 2010) and are thus important to many retailers (Yadav and Pavlou 2014), they are not always the eventual outcome of strategic interest.

Online retailers have strategic interest in the basket value of a transaction, because their profits are directly tied to the shipping costs associated with the purchases (Boatwright, Borle, and Kadane 2003). Often, these costs account for more than half of the firm's operating costs (O'Neill and Chu 2001), marking a difference from traditional offline (brick-and-mortar) stores, for which total sales are the primary focus. Some researchers have investigated purchase quantities (Boatwright et al. 2003), but the factors that affect basket value have not been explored.

Consumers may expend time browsing pages and gathering information before making a purchase. Previous research has shown that consumer online purchase behavior can be better predicted by incorporating browsing characteristics as covariates (Montgomery et al. 2004; Sismeiro and Bucklin 2004) and that browsing characteristics such as page views and visit duration are jointly determined within a browsing session (Bucklin and Sismeiro 2003). Integrating various approaches to modeling the diverse aspects of the browsing process, we conceptualize two distinct stages: (1) the browsing stage, wherein the consumer expends time (visit duration) viewing a number of web pages (page views), which may then result in a purchase decision, conditional on which there is (2) the purchase stage, wherein a basket of certain value is realized.

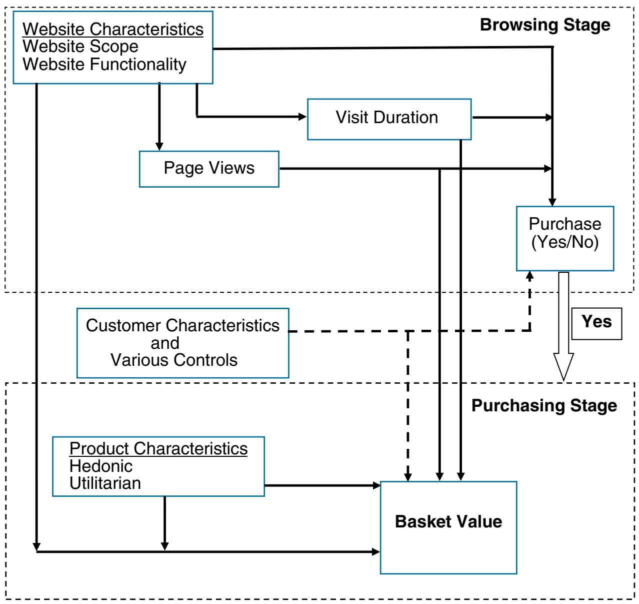
Another important missing link in this setting involves the role of factors that might jointly affect the consumers' online browsing and purchase behavior. Although the role of contextual factors in the traditional retail shopping (Zhang et al. 2014, Inman, Winer, and Ferraro 2009) experience are well investigated (e.g., Grewal et al. 2009), our understanding lags in online retail settings. Although research on click-stream data has explored behavior within a specific session (e.g., Moe 2006), the role of contextual factors across sessions that might jointly influence the overall shopping behavior is an emerging area of inquiry. Because consumers purchase specific products and services from a specific online retailer, we posit that the key contextual factors are the product, or what is being shopped, and website characteristics, or where it is being shopped (Bart et al. 2005; Zhu and Zhang 2010). Therefore, with the current research we seek to understand how product and website heterogeneity influence page views, visit duration, purchase decision, and basket value, independently and jointly, after controlling for consumer heterogeneity.

Figure 1 provides a pictorial representation of our conceptual framework. We capture the staged nature of the browsing and purchase process and evaluate the impact of various drivers that lead to purchase and ultimately basket value. The Web Appendix (see Table WA-1) provides a summary of the main characteristics of our approach, in comparison to extant research on the modeling of online behavior.

The Role of Contextual Factors: Product and Website Heterogeneity

When consumers shop online, various factors influence their purchase experience, including the browsing process and the final purchase decision (Bart et al. 2005; Bucklin and Sismeiro 2003). Conceptually, we focus on the context-specific factors (the what) and the website on which the shopping experience occurs (the where), while acknowledging the role of consumer heterogeneity. From a strategic standpoint, the context-specific factors, and their potential interplay, could be strategic levers for achieving desired firm-level objectives (e.g., Hong, Thong, and Tam 2004; Zhu and Zhang 2010). Therefore, in addition to their main effects we also accommodate the notion that product and website heterogeneity could jointly determine the outcomes of the purchase experience, which constitutes an important contribution of our research.

FIGURE 1
Browsing and Purchasing Model



Notes: Page views, duration, purchase incidence, and basket value are endogenous variables. Purchasing stage is conditional on purchase incidence being equal to 1.

Extant evidence points to a crucial role of product characteristics in determining shopping behavior. For example, drawing on regulatory focus theory (Chernev 2004; Higgins 1997), Chitturi, Raghunathan, and Mahajan (2007) argue that when consumers shop, functional characteristics of the products help achieve prevention goals, which emphasize safety and prevention of negative outcomes, whereas hedonic characteristics do the same for promotion goals, which emphasize on positive outcomes and accomplishments. Kushwaha and Shankar (2013) investigate the moderating role of product characteristics in a multichannel environment and find that monetary value of purchase is higher for single-channel shoppers than multichannel shoppers for utilitarian products, whereas the effect is reversed for hedonic products.

Thus, in an online setting as consumers browse in pursuit of their shopping goals, we reason that their shopping behavior will be influenced by the characteristics of the product categories in which they may be interested. Given that hedonic and utilitarian characteristics invoke promotion and prevention foci (Chernev 2004), respectively, we expect that the product heterogeneity will influence the monetary value of the transaction (e.g., Kushwaha and Shankar 2013). Given that insights on which products are more amenable to online shopping are only emerging (Huang, Lurie, and Mitra 2009), as well as potential interactions between product characteristics and other factors, we do not develop explicit expectations regarding the directions of the main effects of product heterogeneity.

In addition to what is being shopped, we expect that where it is being shopped also influences shopping behavior. As websites are primary online platforms for the shopping experience, their characteristics strongly influence the customer experience and thus the outcomes (Rose et al. 2012). For example, research indicates that elevated levels of advertising lowers visit durations (Danaher et al. 2006) and that design, layout, and content of retailers' websites influence consumer shopping outcomes in general (Galletta et al. 2006; Hong et al. 2004). We focus on two aspects of retailer website heterogeneity: its scope in terms of product variety and its functionality.

Despite a general sense of how retailers differ in their strategic orientation—retailers might offer a narrow or a broad range of product categories—we know of few insights into how this orientation affects consumer behavior in online shopping contexts. Recently, in their study of online search behavior of consumers, Narayanan and Kalyanam (2015) distinguish retailers according to the breadth of their product assortment. Given our context, we build on their approach and conceptualize website scope as the breadth of product categories the online retailer offers. Thus, in our framework websites with a broad scope are generalists, and those with a narrow scope are specialists (e.g., Swaminathan 1995). As dominant broad-scope retailers such as Amazon.com continue to grow and upend retailers with fewer categories, the strategic orientation debate persists, and its impact on how consumers browse online could provide significant insights for firms. All else being equal, as retailers with broader product scope offer an opportunity to fulfill overall shopping needs with better pricing deals (Bell and Lattin 1998), we expect that the basket value will be higher for such retailers.

As consumers navigate the retailer's website, they use multiple website features as they accumulate page views and prolong visit duration. For example, they may visit the FAQ section or the community forum to understand how to obtain additional product information or access specific parts of the website (e.g., Bart et al. 2005). They may also use features that help them navigate better, such as a website maps and design and presentation elements (Hong et al. 2004). The website's functionality can either facilitate or constrain consumers in their browsing path, thereby influencing the eventual shopping outcomes. On the one hand, websites that offer rich functionality might equip the consumers with better tools to gather information and positively influence some browsing outcomes, such as visit duration (Danaher et al. 2006). On the other hand, emerging research on effective website design indicates that richer functionality might create clutter on websites, impede the consumer search process, and negatively influence outcomes (e.g., Galletta et al. 2006). Therefore, we expect functionality to play a mixed role when it comes to influencing shopping outcomes.

Interaction Between the Contextual Factors

Because the retailer's website imposes constraints on how product information will be presented to consumers, we expect that the product's heterogeneity as characterized by its hedonic/ utilitarian characteristics will interact with the website's characteristics in determining shopping outcomes (Huang, Lurie, and Mitra 2009). For example, while purchasing utilitarian products such as printers, shoppers might seek to access product manuals and other information, whereas while shopping for hedonic products such as jewelry, they might want to look at 360-degree views of images or talk to customer representatives. Thus, the extent to which such product-related actions are possible is determined by the website's functionality.

The hedonic or utilitarian nature of a product, in conjunction with the scope of the retailer's website, could influence the value of a consumer's basket. The wider assortment of categories available on the website could keep consumers more engaged (González-Benito and Martos-Partal 2012) and endow them with feelings of abundance, such that they spend more money per transaction for hedonic products, whose purchases are driven more by affect (Khan and Dhar 2006; Richins 2013) than reason. Because these evaluations involve subjective assessments (Bhargave and Montgomery 2013) and hedonic purchases involve variety seeking (Garg, Inman, and Mittal 2005), the availability of a wider variety facilitates greater motivation to purchase and subsequently more spending on purchases. With respect to utilitarian products, consumers are driven primarily by appeals to reason rather than emotion (Chitturi et al. 2007), and with a narrow website scope, carrying limited variety might constrain the consumer and lead to a perception of suboptimal decision making. Utilitarian purchases are prevention focused (Chernev 2004); thus, consumers should prefer to shop at websites that offer greater variety, where they can efficiently gauge the overall choice and make a decision. Therefore, we expect that purchases at websites with broader product scope would also lead to baskets of higher value for utilitarian products.

Due to the varying nature of information sought by consumers while shopping for various products, they might visit different parts of the website and gather information differently (e.g., Huang, Lurie, and Mitra 2009). We propose that a product's hedonic/utilitarian nature could prompt the consumer to look on the website for specific type of information to aid their judgment. Our expectation stems from signaling theory (Benartzi, Michaely, and Thaler 1997). Because hedonic products are inherently more experiential or aesthetic, access to user forums in which social approval can be obtained (Wells, Valacich, and Hess 2011) and holistic judgments qualified through communication can help increase basket value. Moreover, hedonic purchases occur infrequently, are associated with strong emotional appeal involving pleasure (Goldsmith, Cho, and Dhar 2012) and guilt (e.g., Khan and Dhar 2006), and make consumers promotion focused (e.g., Chernev 2004). Therefore, interaction with other users for affirmation through reading forums, the ability to talk to customer service, and so on might help consumers convince themselves and increase eventual basket value because the consumer might feel the need to justify such purchases (Okada 2005). However, when other elements such as navigational features interfere with this consumer experience and create clutter (Galletta et al. 2006), they might feel hindered, which may lead to lower basket value.

In contrast, for utilitarian products, which are more functional and practical, access to additional information such as product manuals, technical specifications that can be accessed on related pages would signal quality and hence, enhance the appeal for the product. As the utilitarian nature of the product increases, certain aspects of the online retailer's website functionality, such as site layout, can be beneficial in linking it to ancillary information appealing to the consumer's reason and thus synchronizing with their prevention focus (Chitturi et al. 2007). However, the presence of communication elements might cause redundancy of information given that utilitarian purchases are often viewed as a chore. Consumers often have experience buying such products in their everyday life (Okada 2005), need little affirmation to purchase, and are prevention focused during the experience (Kushwaha and Shankar 2013).

In summary, with respect to the interaction between product heterogeneity and website scope we expect basket value to be greater for both hedonic and utilitarian products when shopping at websites with broader scope. For the interaction between product heterogeneity and website's functionality, we expect mixed effects, such that purchases of hedonic products might result in greater basket value due to certain aspects of functionality that facilitate communication and interaction or lower basket value due to functionality characteristics that facilitate navigation and layout. Conversely, for utilitarian products, we expect the basket value to be lower due to aspects related to communication and interaction and greater due to functionality aspects such as navigation and layout.

Controlling for Consumer Heterogeneity

The impact of consumer heterogeneity on purchase behavior is well documented in marketing literature (e.g., Chandukala et al. 2011; Fader and Hardie 1996), which is important because consumers who are likely to exhibit a specific, desirable reaction to marketing stimuli then can be profiled to create an effective targeting strategy (Li and Kannan 2014; Young, DeSarbo, and Morwitz 1998). Accordingly, we control for the effects of key demographics—household income, size, race, census region, and Internet connection speed—in addition to customers' overall propensity to shop online.

In summary, we propose a model of online consumer browsing and purchasing behavior to explore the effects of browsing behavior (page views and duration) and product and website characteristics on the eventual outcome (i.e., basket value conditional on consumer's decision to purchase after controlling for consumer heterogeneity). To investigate these phenomena, we need data about online browsing sessions with transactions across (1) product categories, (2) online retailer websites, and (3) consumers. We thus elaborate on the characteristics of the data that we acquired to investigate these questions and develop our proposed model next.

Methodology

Data

We assembled a large and unique data set from multiple sources. In particular, we obtained transaction-level household panel data from the ComScore Web Behavior Panel for 2011 (ComScore 2011). The ComScore Web Behavior data include the browsing and buying behaviors of online users from the United States. We obtained the panel from a random sampling of more than 2 million Internet users whose online activity ComScore has explicit permission to access and capture. The data capture, at the individual household level, information about every website visited, page views, and the duration of visit at each website during a browsing session. Furthermore, these data include session- and transaction-level information, such that we can capture the heterogeneity in multiple variables of interest. ComScore data also have been used widely in prior marketing research (e.g., Johnson et al. 2004; Zhu and Zhang 2010).

We randomly selected 2,000 households from the overall sample and retained the entire browsing and purchase history for these households. This resulted in a total of 773,262 individual browsing sessions, of which 9,664 resulted in purchases across 43 product categories and 385 online retailers. Note that a browsing session is specific to a website (according to ComScore data definitions). Thus, for every session, the consumer and website are unique, but multiple product categories could be in the basket when a purchase occurs. Because we intend to analyze a specific browsing session, we operationalized the outcome variables at the session level. Page views (PVIEW) is the number of pages viewed in that session, and visit duration (DUR) is the length of time in minutes of that session. Similarly, basket value (BVAL) is the dollar value of the basket that the consumer ultimately purchased in a given session.[1]

To calibrate products on their hedonic/utilitarian characteristics, we augmented the ComScore data with an online study, using respondents from Amazon's Mechanical Turk. We used the scale developed by Voss, Spangenberg, and Grohmann (2003) for scoring hedonic and utilitarian characteristics of product categories. One hundred twenty participants provided ratings on five-item scales for the two attributes. We conducted a confirmatory factor analysis to ensure that the items loaded on the two theoretical factors. We then retained the factor scores for hedonic (hedon) and utilitarian (util) characteristics of the product categories in the study.[2]

We accounted for website heterogeneity with two measures: scope of the retailer's website in terms of product variety offered and the functionality of the website. To categorize the retailer's website on the basis of the scope of product variety (Narayanan and Kalyanam 2015), we relied on basic definitions: retailers that sell products across different categories have broad scope, whereas retailers that sell limited variety have narrow scope. To be consistent in our coding scheme, we used the categories defined by ComScore. We hired two independent raters to categorize each retailer in our sample as having broad or narrow website scope in terms of product category variety. The raters visited each website in the sample, browsed the offerings, and then coded the website, while remaining cognizant of the product category definition provided by ComScore. To validate these findings, we also applied a second method, in which we captured the number of product categories in which transactions occurred for a given retailer across the entire sample. We categorized the retailers as having broad scope if their websites provided more than five categories and as having narrow scope if they offered fewer than five categories. We used this categorization to code a dichotomous variable, Website_Broad, which took a value of 1 if the retailer's scope was broad and 0 if it was narrow. The categorization developed by the raters in the first step correlated well (.73) with the dichotomous measure, lending confidence to our categorization scheme.

[1]Multiple baskets could be purchased within a session on the same website, so we averaged these basket values when they occurred; 65% of purchases were single basket, 21% were two baskets, and 13% were three or more.

[2]When multiple product categories were in a basket, we averaged them across categories.

To develop the web functionality score for each retailer's website, we replicated the procedure used by Danaher et al. (2006) who built on prior research on website functionality (Ghose and Dou 1998), and developed the measure, which controls for website-specific functionality effects. Specifically, three independent raters coded whether the 385 websites in the database featured each of the 19 attributes used to measure a website's functionality. Sample attributes included availability of registration, chat room, help section, and product information, among others. Their interrater reliability was .87; therefore, we used the average across the three raters to obtain the final measure. Because of the rich variety of features used to develop the measure, we further investigated the scale to determine whether we could develop a more nuanced view of the construct. To do so, we conducted an exploratory factor analysis on the 19 items on which the websites were scored and found two underlying dimensions, which we labeled communication (CF) and navigational (NF) functionality. Communication functionality captures the extent to which the website offers communication-oriented features (e.g., e-mail id, chat rooms, message boards), whereas navigational functionality captures the extent to which the website facilitates browsing through features such as access to the website maps, content, layout, and updates. We believe that capturing these aspects of website characteristics provides a more nuanced understanding of consumer browsing and purchase behavior. We estimated the factor scores and used them as measures for the website's communication and navigational functionality.

We provide summary statistics about the characteristics of the top 20 product categories, ranked by total number of transactions, in Table 1 and summary statistics about the characteristics of the top 20 online retailers, ordered by total number of transactions, in Table 2. These tables highlight the complex variation in browsing patterns.

We control for some possible alternative explanations that might arise due to variation in consumer characteristics. Two dummy variables accounted for the racial background of the head of the household, with Caucasian as the baseline (African American, race_af, and Asian, race_as, are the other two categories) using the categorization provided by ComScore. We coded annual household income as low, medium, or high, using ComScore's categorization (low: <$35,000; medium: $35,000–$75,000; high: >$75,000). For household size, we coded three categories, again using ComScore's categorization (small: 1–2 household members; medium: 3–5 household members; large: 5 or more household members). The low category was the baseline category for income, and the small category provided the baseline for household size. We also included dummies to account for variation arising due to the census region in which the consumer resided (four regions: northeast, north central, south, and west, with northeast as the baseline). Overall, the head of the household was Caucasian in 75% of the households, African American in 22%, and Asian in 4%. In terms of income heterogeneity, 29% of the households were low income, 39% earned medium income, and 32% were in the high category. Furthermore, 39% were small households, 38% were medium, and 23% were large households. To provide a more detailed picture of the distribution of households in the sample, we created Table WA-2 (in the Web Appendix) in which we break down the households across racial backgrounds, income categories, and household sizes.

In addition to the demographics, we included a set of control variables related to the consumer's browsing behavior as well. First, we used the speed of the Internet connection, as provided by ComScore (Connection_Speed), as a dichotomous variable taking a value of 1 if the household had a broadband connection and 0 otherwise. Because the three focal outcomes may depend on the consumer's intrinsic shopping propensity, we included a control variable Dom_Variety equivalent to the average number of unique domains the consumer visited up until the month of the focal transaction to measure extent of the consumer's shopping propensity. We also included dummies to account for variation across calendar months and various other controls (e.g., price dispersion, domain variety), which we discuss in more detail in the following section. We present the descriptive statistics and bivariate correlations for the variables in the model estimation in Table 3 and additional details about the characteristics of the sample in Web Appendix (Figures WA-1 and WA-2).

Modeling Approach

Our primary focus in this research is to understand the impact of various product- and website- related factors on basket value. However, basket value for a browsing session can only exist if a consumer has made a purchase in a browsing session. To develop the most appropriate model, we need to account for purchase incidence, which addresses the selection issue pertaining to consumer browsing and purchase decisions. Purchase incidence takes a value of 1 if the consumer makes a purchase in a given session and 0 otherwise. Therefore, by incorporating the selection process we account for potential bias in the regression parameters in the basket value equation, which exists only when there was a purchase made.

To account for this selection bias, we employ the Heckman selection model (1979), which has found other applications in marketing literature (e.g., Bucklin and Sismeiro 2003; Ying, Feinberg, and Wedel 2006), with a few important modifications that enhance the methodological contribution of our research. First, the selection model accounts for purchase incidence (i.e., selection issue), conditional on which the basket value (BVAL) is observed. Second, purchase incidence is specified as a function of browsing characteristics, including page views (PVIEW) and visit duration (DUR) that are both endogenously determined in addition to the consumer's past purchase behavior and other consumer and website characteristics. We then jointly estimate the selection and the subsequent equations, conditional on the selection criterion, in a single step. Our approach is similar to other applications that involve a system of equations featuring selection (Petersen and Kumar 2009).

The purchase incidence (selection) model is specified as follows:

$$\text{Purchase}_{ij} = \begin{cases} 1 & \text{if } \text{Purchase}^*_{ij} \geq 0 \\ 0 & \text{otherwise} \end{cases}, \tag{1}$$

where Purchase_{ij} is 1 if the latent variable Purchase^*_{ij} is greater than or equal to zero for consumer i in a browsing

TABLE 1
Summary Characteristics of Top 20 Product Categories (Ordered by Number of Transactions)

Product Category	Number of Transactions in Product Category	Number of Unique Households that Purchased in the Product Category	Hedonic Value	Utilitarian Value	Average Communication Functionality	Average Navigational Functionality	Average Page Views	Average Visit Duration	Average Basket Value
Food & Beverage	1,758	371	81.05	97.82	44.97	15.98	8.67	11.78	24.64
Apparel	1,569	443	65.52	88.04	70.11	13.08	35.13	24.86	75.72
Books & Magazines	876	310	100.00	65.90	48.96	100	28.49	19.35	38.24
Health & Beauty	777	243	49.80	58.22	50.81	20.76	31.57	22.04	50.40
Movies & Videos	467	205	90.44	50.31	84.68	91.25	26.53	16.84	45.24
Shoes	348	217	65.22	100.00	78.93	25.08	28.75	21.74	75.85
Toys & Games	262	154	24.55	75.97	53.84	26.33	31.17	25.35	69.40
Arts, Crafts & Party Supplies	249	89	38.51	80.42	54.80	15.77	32.89	28.94	66.15
Mobile Phones	208	101	60.68	84.93	80.95	58.40	18.09	15.99	38.29
Other Computer Supplies	200	119	48.33	41.52	86.38	27.90	22.20	15.63	74.69
Accessories	199	137	67.51	28.70	63.44	18.56	35.48	26.28	65.00
Gift Certificates & Coupons	169	99	74.14	94.73	38.73	68.89	14.52	12.29	24.48
Printers, Monitors & Peripherals	165	98	43.47	44.81	88.88	32.68	25.63	18.41	126.67
Bed & Bath	163	99	43.98	78.52	69.37	13.05	30.71	24.89	80.05
Other Electronics & Supplies	162	104	46.33	68.96	96.71	45.85	24.74	19.72	47.42
Kitchen & Dining	158	98	43.61	79.09	70.71	35.57	28.15	22.30	79.89
Sport & Fitness	157	115	70.80	58.86	88.87	35.32	25.67	21.18	87.68
Computer Software	151	92	52.71	55.89	26.26	14.45	16.89	13.71	72.84
Other Home and & Living Items	142	88	35.28	50.23	73.02	17.63	30.48	21.53	91.32

Notes: Average communication and navigational functionality are the averages score across all websites from which the product category was purchased. We transformed the website functionality, hedonic, and utilitarian scores to a 0–100 scale for ease of presentation.

TABLE 2
Summary Characteristics of Top 20 Websites (Ordered by Number of Transactions)

	Number of Transactions at the Retailer's Website	Number of Unique Households that Browsed at the Retailer's Website	Average Hedonic Value	Average Utilitarian Value	Website Scope	Communication Functionality	Navigational Functionality	Average Page Views	Average Visit Duration	Average Basket Value
Online Retailer1	2,476	1,486	52.19	53.86	Broad	40.47	56.00	9.69	7.46	50.0
Online Retailer2	615	228	74.77	55.23	Narrow	100.00	80.98	4.60	9.00	21.5
Online Retailer3	564	264	60.14	53.06	Narrow	100.00	80.98	7.27	8.53	21.3
Online Retailer4	413	206	74.63	55.11	Narrow	100.00	80.98	10.34	9.32	21.8
Online Retailer5	345	1,197	74.77	55.23	Broad	44.58	49.11	8.36	6.41	66.5
Online Retailer6	293	466	69.24	12.84	Narrow	49.24	51.38	6.52	6.64	61.3
Online Retailer7	247	76	55.95	59.91	Narrow	42.83	48.31	27.83	11.51	33.2
Online Retailer8	182	404	69.24	12.84	Broad	46.41	51.43	15.16	5.60	63.8
Online Retailer9	156	349	61.73	61.66	Narrow	44.29	48.44	11.96	6.54	87.9
Online Retailer10	154	350	57.17	13.24	Narrow	44.57	46.36	13.79	7.58	102.1
Online Retailer11	132	608	66.79	61.17	General	41.07	49.78	11.41	6.75	81.3
Online Retailer12	118	120	45.57	56.00	Narrow	70.22	47.70	14.02	11.55	90.9
Online Retailer13	105	165	28.44	71.92	Narrow	45.30	46.13	12.40	6.90	78.9
Online Retailer14	103	917	62.32	66.52	Narrow	43.84	48.71	7.75	5.44	80.4
Online Retailer15	100	168	60.61	57.28	Narrow	44.83	51.36	11.04	7.69	67.1
Online Retailer16	94	176	60.11	61.48	Narrow	57.91	9.86	5.17	7.45	64.4
Online Retailer17	87	314	45.53	42.67	Broad	34.92	60.09	10.07	6.35	25.9
Online Retailer18	85	47	60.24	37.48	Narrow	43.97	47.14	17.38	11.05	86.8
Online Retailer19	83	502	62.94	66.41	Broad	46.05	54.70	10.88	5.23	125.7
Online Retailer20	79	1,325	69.24	12.84	Broad	28.59	65.38	12.26	8.74	36.8

Notes: The names of the retailers are not listed, in accordance with the ComScore licensing agreement. The average hedonic and utilitarian values for a website are the average scores across all the product categories purchased from that website. Summary statistics for the 385 retailers are available from the authors on request. We transformed the website functionality, hedonic, and utilitarian scores to a 0–100 scale for ease of presentation.

TABLE 3
Summary Statistics

Variables	Mean	SD	(1)	(2)	(3)	(4)	(5)	(6)
1. Log_PageViews (PVIEW)	1.95	.87	1.00					
2. Log_Duration (DUR)	2.20	1.26	.47	1.00				
3. Log_BasketValue (BVAL)	3.97	1.04	.31	.25	1.00			
4. Hedonic (hedon)	66.47	20.64	.13	.10	.26	1.00		
5. Utilitarian (util)	73.92	21.23	–.23	–.10	–.17	–.08	1.00	
6. Communication Functionality (CF)	79.09	38.79	–.11	.15	–.05	.17	–.11	1.00
7. Navigational Functionality (NF)	90.38	4.36	–.01	–.01	.14	–.11	–.02	–.16

Notes: Number of browsing sessions = 773,262. All correlations above .01 are significant at $p < .01$. We present the mean and standard deviation reported for product characteristics (hedonic and utilitarian) and website functionality (communication and navigational) as values after transforming the measures to a 0–100 scale.

session j, j = 1, …, N_i. The latent variable Purchase^*_{ij} is modeled as follows:

$$(2)\quad \text{Purchase}^*_{ij} = \alpha_0 + \alpha_1(\text{PVIEW}_{ij}) + \alpha_2(\text{DUR}_{ij}) + \alpha_3(\text{Past}_{\text{purc}_i}) + \alpha_4(X^W_{ij}) + \alpha_5(X^C_{ij}) + \alpha_6(\text{Dom\_Variety}_i) + \delta_{1i} + U_{ij},$$

where X^W captures website characteristics and X^C captures consumer characteristics and time dummies. In particular, X^W = [Website Scope, Communication Functionality and Navigational Functionality], and X^C = [race dummies, household income, household size, census region dummies, connection_speed, monthly dummies]. Consistent with prior research (Bucklin and Sismeiro 2003; Danaher, Mullarkey, and Essegaier 2006), we use the log-transformation of page views (PVIEW) and visit duration (DUR). This approach accommodates the right-skewed nature of the variables without complicating the model structure. The variable $\text{Past}_{\text{purc}_i}$ in Equation 2 is the cumulative number of purchases made by the consumer, and Dom_Variety_i is the number of unique domains visited by the consumer up until the month in which the focal transaction occurs. δ_{1i} is the consumer-level random effect, and U_{ij} is the residual error.

Furthermore, number of pages viewed and duration of time spent browsing could be endogenous to a given browsing session. Therefore, we account for this endogeneity using the following equations:

$$(3)\quad \text{PVIEW}_{ij} = \gamma_{10} + \gamma_{11}(\text{Past}_{\text{Pview}_i}) + \gamma_{12}(X^W_{ij}) + \gamma_{13}(X^C_i) + \gamma_{14}(\text{Dom\_Variety}_i) + \gamma_{15}(\text{Dom\_Views}_{ij}) + \delta_{2i} + V_{1ij},$$

$$(4)\quad \text{DUR}_{ij} = \gamma_{20} + \gamma_{21}(\text{Past}_{\text{Dur}_i}) + \gamma_{22}(X^W_{ij}) + \gamma_{23}(X^C_i) + \gamma_{24}(\text{Dom\_Variety}_i) + \gamma_{25}(\text{Dom\_Dur}_{ij}) + \delta_{3i} + V_{2ij},$$

The variable $\text{Past}_{\text{Pview}_i}$ in Equation 3 is the moving average of page views gathered by the consumer, and Dom_ Views_{ij} is the average monthly views gathered by the domain, up to, but not including, the focal transaction and result in the exclusion restrictions pertaining to the page views equation. Similarly, the variable $\text{Past}_{\text{Dur}_i}$ in Equation 4 is the moving average of visit duration for the consumer, and Dom_Dur_{ij} is the average duration gathered by the domain, up to, but not including, the focal transaction and result in the exclusion restrictions pertaining to the visit duration equation. δ_{2i} and δ_{3i} are the consumer-level random effects, and V_{1ij} and V_{2ij} are the residual errors.[3]

If the customer makes a purchase (i.e., the selection occurs) the total basket value is observed; otherwise it is not observed. Therefore, we model the basket value conditional on the occurrence of purchase as follows. Let BVAL_{ik} denote the log of the total basket value for consumer i during purchasing session k, k = 1, ..., M_i,

$$(5)\quad \text{BVAL}_{ik} = \beta_0 + \beta_1(\text{PVIEW}_{ik}) + \beta_2(\text{DUR}_{ik}) + \beta_3(X^W_{ik}) + \beta_4(X^P_{ik}) + \beta_5(X^W_{ik} \times X^P_{ik}) + \beta_6(\text{Price\_Disp}_{ik}) + \beta_7(X^C_i) + \beta_8(\text{Past}_{\text{Bval}_i}) + \beta_9(\text{Dom\_Variety}_i) + \beta_9(\text{Basket\_Variety}_{ik}) + \delta_{4i} + \varepsilon_{ik} \text{ if } \text{Purchase}_{ij} = 1.$$

As more than one product category may have been purchased in a basket, we control for this using three variables—Price_Disp_{ik} captures the dispersion of prices in the basket (measured as variance), $\text{Basket\_Variety}_{ik}$ is a measure of the number of product categories in the basket. δ_{4i} is the consumer-level random effect, and ε_{ik} is the residual error.

And finally, factors not observed by the researchers but known to consumers could cause common shocks that affect customers' browsing and purchase behavior. To account for these unobserved common shocks, we specify the following:

$$(6)\quad \begin{bmatrix} \delta_{1i} \\ \delta_{2i} \\ \delta_{3i} \\ \delta_{4i} \end{bmatrix} = \text{MVN}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \Sigma_\delta\right) \quad \text{and} \quad \begin{bmatrix} U_{1i} \\ V_{1ij} \\ V_{2ij} \\ \varepsilon_{ik} \end{bmatrix} = \text{MVN}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \Sigma_\varepsilon\right),$$

[3]We verify the validity of these instruments as part of the empirical analysis.

where $\delta_{1i}, ..., \delta_{4i}$ are consumer-specific random effects for Equations 2–5 that may correlate with one another (and Σ_δ is the variance–covariance matrix). Furthermore, $U_{ij}, ..., \varepsilon_{ik}$ are equation-specific errors (Σ_ε is the variance–covariance matrix; $Var(U_{ij}) = 1$ for facilitating the probit formulation).

Therefore, we propose a system of equations with correlated error terms to account for potential endogeneity among the dependent variables (e.g., Petersen and Kumar 2009). We consider a range of factors that could drive purchase, page views, and visit duration taken from prior research findings (Sismeiro and Bucklin 2004; Yadav and Pavlou 2014). The model also accounts for the unbalanced nature of the data, because the number of purchases for consumers r and p could differ (M_r and M_p).

In summary, we jointly model purchase incidence (selection stage) and the value of the basket conditional on the selection stage, in addition to treating page views and duration as endogenous and specifying appropriate instruments, in addition to letting the equation-level errors to covary. Importantly, we also incorporate heterogeneity at the product, website, and consumer level, in addition to accounting for unobserved heterogeneity at the consumer level. As the model incorporates variation at different levels, we use a multivariate mixed-effects model (Lindstrom and Bates 1990; Verbeke and Lesaffre 1996) that is appropriate for the context and has found applications in the marketing context (e.g., Krasnikov and Jayachandran 2008; Landwehr, Wentzel, and Herrmann 2013). However, we build on the approach's standard specification by developing a system of equations to incorporate selection and the presence of endogenous regressors in both the purchase incidence and basket value equations. We have a Type II Tobit model in which there is a selection stage followed by a transaction stage, but some regressors in both the selection state and the transaction stage are endogenous (page views and duration).

To accommodate the various features of the model, we used the Conditional Mixed Process routine (command CMP) in Stata (Roodman 2011) to estimate the model. At its core, CMP is a seemingly unrelated regression framework that is flexible enough to incorporate features such as correlated errors and random effects across equations. Furthermore, it allows regressors to be endogenous as well. The estimation uses maximum-likelihood and the random effects are simulated using draws following a Halton sequence by relying on the DFP search algorithm (for a more detailed review see Drukker and Gates 2006). Subsequently, we also elaborate on alternative specifications of the proposed model to ensure the validity of our findings.

Results

We estimated alternative models for the purpose of benchmarking. We created the alternative set by beginning with a base model with only control variables and then added different aspects of the proposed model in a sequential manner. Thus, we estimated Model 1 with only control variables (consumer characteristics, browsing behavior and time dummies) as explanatory variables, and in Model 2 we added the main effects of the contextual variables, product and website characteristics. Models 1 and 2 also incorporated consumer-level random effects. In Model 3, we included the main effects and interactions between the contextual factors but did not include the consumer-level random effects. Model 4 is the proposed model in which we added the consumer-level random effects to the model that included main and interaction effects in addition to all the control variables. Model 4 outperformed other models on standard information criteria (log-likelihood, Akaike information criterion, and Bayesian information criterion) used to assess model fit. We provide model fit criteria and details regarding model comparison in Table 4. Next, we present the estimates obtained from the proposed model (Model 4).

Results of the Browsing Model: Purchase Incidence, Page Views, and Visit Duration

Table 5 presents the results pertaining to the browsing stage leading up to purchase incidence (Equations 2–4), and then Table 6 displays the results for the final outcome variable, basket value (focusing on Equation 5),[4] in greater detail. Focusing on the important effects in the selection model—purchase incidence (Equation 2)—we find that consumer's past browsing behavior matters. Past purchase incidence, has a significant positive impact on the probability of purchase (.112, $p < .01$), implying that consumers who purchased more frequently in the past, all else being equal, have a greater likelihood of purchase in a given session. Consistent with extant research (e.g., Danaher 2007; Danaher et al. 2006), which has conceptualized page views and duration as interest and effort of the consumer, we find that they are positively associated with purchase probability (.239, $p < .01$; .319, $p < .01$, respectively). With respect to website characteristics, the likelihood of purchase is higher at websites with broad scope as indicated by the positive effect for the website scope dummy (.220, $p < .01$). It seems that the flexibility in shopping at a website that offers multiple product categories leads to better conversion, which perhaps explains why other retailers view broad scope retailers as a strategic threat (Brandt 2011). Interestingly, we find evidence that communication functionality is negatively associated with likelihood of purchase (–.281, $p < .01$). Although this result seems counterintuitive, it is in tune with recent research on website design that indicates that website elements that are perceived to be irrelevant to the browsing context could impede the shopping process (e.g., Cuddihy and Spyridakis 2012; Wells, Valacich, and Hess 2011). We reason that as consumers have become relatively comfortable in using websites, on average, they increasingly feel these common features are a hindrance to their experience and do not appreciate the clutter caused by the various communication elements.

For page views (Equation 3),[5] we find that the consumer's past page visit behavior positively influences current page

[4]Note that the results were obtained in joint estimation but are presented in separate tables for ease of exposition.

[5]Following an instrumental variable regression using two-stage least squares, we tested the null hypothesis that the instruments are weak, using the minimum eigenvalue statistic proposed by Cragg and Donald (1993). The null was rejected at a 5% relative bias (test statistic: 48.76, critical value = 11.04), rejecting the null that the instruments were weak. Next, the Sargan ($\chi^2_2 = 3.49$, $p = .17$), Bassman ($\chi^2_2 = 1.74$, $p = .18$), and Wooldridge's robust score ($\chi^2_2 = 3.99$, $p = .14$) tests provided strong and converging evidence that the instruments were valid.

TABLE 4
Model Comparison

Model	Main Effects of Theoretical Variables (Model Feature)	Interactions (Model Feature)	Consumer-Level Random Effect (Model Feature)	Log-Likelihood	Akaike Information Criterion	Bayesian Information Criterion
M1	No	No	No	–2,135,713	4,271,652	4,272,959
M2	Yes	No	No	–2,109,778	4,219,810	4,221,278
M3	Yes	Yes	No	–2,109,679	4,219,624	4,221,161
M4	**Yes**	**Yes**	**Yes**	**–2,103,498**	**4,207,269**	**4,208,841**

Notes: The model with the best fit is indicated in boldface.

views (.028, $p < .01$). However, when it comes to the website's characteristics, websites with broad scope have fewer page views than those with narrow scope, as indicated by the negative and significant effect of the website scope dummy (–.167, $p < .01$). A potential explanation is that websites with broad scope optimize for product breadth and thus enable consumers to investigate various alternatives on a given page without having to navigate a lot of pages (e.g., Amazon.com provides information about various options and alternatives for products on a single page, thus reducing the need to navigate to multiple pages, thereby preventing consumers from changing their minds and resulting in higher conversion rate). With respect to website functionality, we find that both communication (–.021, $p < .01$) and navigational functionality are negatively associated with page views (–.041, $p < .01$). The data indicate that, all else being equal, consumers do not visit many web pages at functionally rich websites whose features are likely to keep them preoccupied and provide the required information on a given page.

With respect to visit duration (Equation 4), the consumer's past behavior continues to be an accurate predictor: we find that past visit duration has a positive and significant effect on a given session's visit duration (.023, $p < .01$). With respect to the website characteristics, we find that websites with broad scope have longer visit durations, all else being equal, as indicated by the positive and significant

TABLE 5
Results of the Browsing Stage

Variable Category	Variables	Probability of Purchase[a]	Page Views	Visit Duration
	Intercept	–3.341*** (.072)	1.560*** (.017)	1.251*** (.019)
Browsing behavior	Page views	.239** (.021)	—	—
	Duration	.319** (.022)	—	—
Website characteristics	Website scope	.220*** (.014)	–.167*** (.003)	.067*** (.004)
	Communication functionality	–.281*** (.006)	–.021*** (.006)	.036*** (.001)
	Navigational functionality	.006 (.009)	–.041*** (.004)	.040*** (.005)
	Domain variety	–.002*** (.000)	.000*** (.000)	.000*** (.000)
Control variables	Past incidence	.112*** (.001)	—	—
	Past page views	—	.028*** (.001)	—
	Past duration	—	—	.023*** (.000)
	Domain views	—	–.000*** (.000)	—
	Domain duration	—	—	.000*** (.000)
	Medium household size	–.078*** (.014)	–.001 (.004)	.007** (.004)
	Large household size	–.134*** (.016)	–.003 (.003)	–.003 (.004)
	Medium household income	.038*** (.015)	.028*** (.003)	–.024*** (.004)
	Large household income	.057*** (.016)	.022*** (.004)	–.017*** (.004)
	African American	–.232*** (.017)	.002 (.004)	.079*** (.005)
	Asian	–.098*** (.036)	.035 (.006)	–.016** (.009)
	North central	–.037* (.019)	.001 (.005)	–.011** (.005)
	South	–.024* (.017)	–.003 (.003)	–.009*** (.004)
	West	–.054*** (.020)	–.006 (.007)	–.012* (.007)
	Connection speed	.126*** (.049)	.011 (.016)	–.046*** (.018)
	Time dummies	Included	Included	Included

*$p < .10$.
**$p < .05$.
***$p < .01$.
[a]Effects of product characteristics are not observed for transactions with no purchase, so these effects cannot be estimated for probability of purchase.
Notes: Values in Tables 5 and 6 were estimated jointly. They are presented in separate tables for ease of presentation.

TABLE 6
Results of the Conditional Purchase Stage

		Basket Value
Intercept		3.420*** (.082)
Browsing behavior	Page views	.168*** (.019)
	Duration	−.037*** (.017)
Product characteristics	Hedonic	−.243*** (.028)
	Utilitarian	−.109*** (.019)
Website characteristics	Website scope (WS)	.111*** (.028)
	Communication functionality (CF)	−.166*** (.047)
	Navigational functionality (NF)	.292*** (.022)
Interaction between product and website characteristics	WS × Hedonic	.160*** (.035)
	WS × Utilitarian	.103*** (.028)
	CF × Hedonic	.209*** (.041)
	CF × Utilitarian	−.069 (.055)
	NF × Hedonic	−.148*** (.018)
	NF × Utilitarian	.123*** (.025)
Control variables	Price dispersion	.000*** (.000)
	Basket variety	.197*** (.013)
	Domain variety	−.002*** (.000)
	Past page views	—
	Past duration	—
	Past purchases	.015*** (.001)
	Customer characteristics	Included
	Time dummies	Included

***$p < .01$.
Notes: Values in Tables 5 and 6 were estimated jointly. We display them in separate tables for ease of presentation. Parameter estimates of customer characteristics are presented in Table WA-3 in the Web Appendix.

effect of the website scope dummy (.067, $p < .01$). Retailers with broad scope provide opportunities to explore multiple categories in any given visit and a convenient one-stop shop to explore, even while not purchasing, which perhaps results in longer visits. Interestingly, a website's functionality in terms of communication (.036, $p < .01$) and navigational characteristics (.040, $p < .01$) are both positively associated with duration. As functionally rich websites provide avenues to explore, gather and collect information (i.e., technical details, customer reviews) along with information about related products, they realize higher visit durations.

In summary, consumer's past behavior emerges as a consistent predictor of browsing and purchasing behavior, highlighting the importance of focusing on retaining loyal consumers even in the digital realm (Shankar, Smith, and Rangaswamy 2003). This finding is consistent with others in the domain on the importance of loyalty and shopping partly due to trust in the platform (González-Benito and Martos-Partal 2012). When it comes to website characteristics, we find that websites with broad scope have a greater probability of purchase and longer visit durations, but fewer page views, primarily due to the one-stop shopping experience. With respect to the website's functionality, as expected we find mixed effects: the communication functionality decreases purchase likelihood and page views, but it improves visit duration; in contrast, navigational functionality decreases page views but improves visit duration.

Results of the Conditional Purchase Model: Basket Value

Table 6 presents the results of the conditional purchase model (effects in Equation 5). Page views are positively associated (.168, $p < .01$), whereas visit duration is negatively associated with basket value (−.037, $p < .01$). All else being equal, it seems that higher basket value is realized when consumers view more page views and quickly make the purchase decision. This is consistent with prior findings that measures of customer engagement at websites, including page views and visit duration, influence the retailer's profit goals (Montgomery et al. 2004), but they also indicate that there might be tradeoffs in optimizing websites for both these outcomes. For websites that derive revenues from advertising (influences by visits) and e-commerce, our results indicate that they stand to gain on both fronts.

We affirmed interaction effects between product characteristics (hedonic and utilitarian) and website characteristics (website scope, communication functionality and navigational functionality). The main effect of both hedonic and utilitarian characteristics is negative on basket value (−.243, $p < .01$; −.109, $p < .01$). The main effect of website scope is positive, indicating that retailers with broad scope realize higher basket value (.111, $p < .01$), whereas the main effects of communication and navigational functionalities on basket value are negative (−.166, $p < .01$) and positive (.292, $p < .01$), respectively. We now interpret how these main effects vary depending on the hedonic/utilitarian nature of the products in the basket.

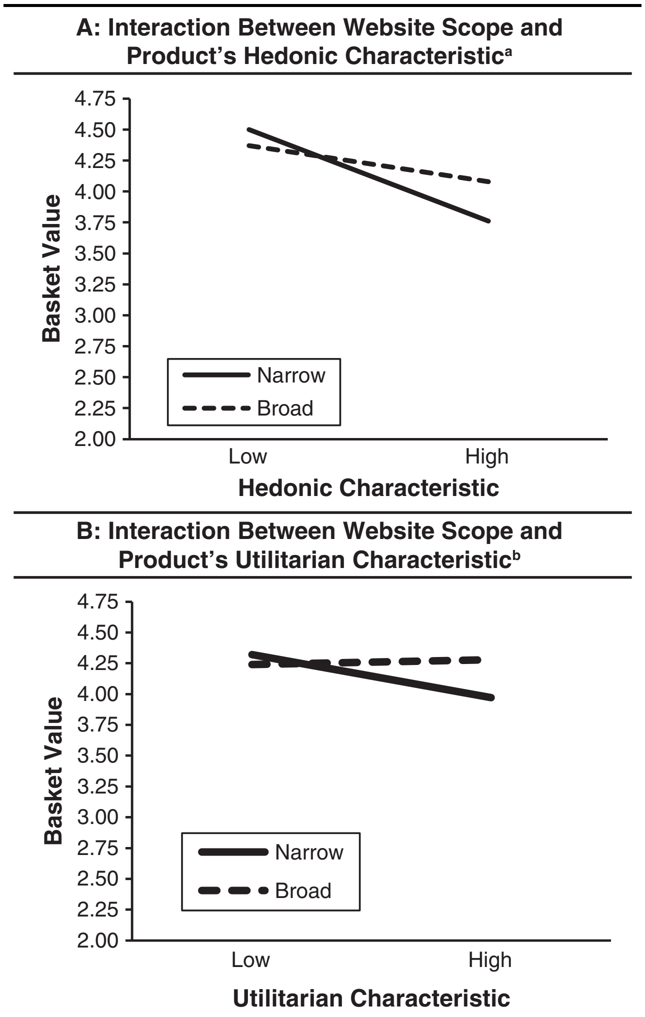
The interaction between website scope and hedonic nature of the basket is positive and significant (.160, $p < .01$). The main effect of website scope dummy is also positive, indicating websites with broad scope on average realize a higher basket value than those with narrow scope; thus, the positive interaction implies that when it comes to hedonic purchases, this difference in basket value is magnified. We observe a similar pattern for the interaction between website scope and utilitarian characteristics of the products purchased (.103, $p < .01$). Thus, overall it seems that retailers with broad website scope can benefit more than retailers with narrow scope if they can highlight the hedonic or utilitarian nature of their products. The wider range of product categories that broad-scope websites carry offers consumers better choices in terms of fulfilling their shopping requirements as the benefits offered by their product categories in terms of hedonic/utilitarian characteristics increase (González-Benito and Martos-Partal 2012). To illustrate further, we developed graphical plots for the significant interaction effects involving the website scope dummy in the basket value equation: with the hedonic (see Figure 2, Panel A), and utilitarian nature of the purchase (see Figure 2, Panel B).[6]

The positive interaction (.209, $p < .01$) of communication functionality with the hedonic nature of the products purchased implies that as the hedonic nature increases, the negative association between communication functionality and basket value weakens. It seems that when consumers can engage in dialogue with others on the website or with the support team, these interactions seem to influence their basket positively for hedonic products. Indeed, hedonic pursuits are emotion driven (e.g., Khan and Dhar 2006) whereas utilitarian are driven more by need and reason (e.g., Chitturi, Raghunathan, and Mahajan 2008); therefore, it is likely that talking and engaging with others is a viable strategy to convince consumers to increase their spending when it comes to hedonic products but not for utilitarian products. This finding is consistent with other research that shows that consumers often seek self-affirmation when it comes to hedonic consumption (Townsend and Sood 2012), and the website's communication functionality facilitates this behavior.

We find that for navigational functionality, the negative interaction with hedonic characteristic (–.148, $p < .01$) implies that as the hedonic nature of the products purchased increases, the positive association between navigational functionality and basket value weakens. One potential explanation for this finding is that because navigational functionality is more focused on getting the consumer to browse and shop the various aspects of the website (e.g., website map, updates), it seems to distract the consumer shopping for hedonic products, which often tend to be expensive and are purchased for emotional reasons. They ultimately obtain more information, but this results in lower basket value. The positive interaction (.123, $p < .01$) of navigational functionality with the utilitarian characteristic implies that as the utilitarian nature of the products purchased increases, the positive association between navigational

[6]We used the ±2 standard deviations of the continuous variables to obtain the high and low levels of the moderator.

FIGURE 2
Interaction Plots Involving Website Scope Dummy for Basket Value



[a]The interaction between website scope and the hedonic nature of the purchase is positive and significant. In Panel A, the plot indicates that as a product's hedonic nature increased, basket value decreased, but the decrease was lower at websites with broad scope. Website scope mitigated the negative effect of the hedonic characteristic.
[b]The interaction between website scope and the utilitarian nature of the purchase is positive and significant. In Panel B, the plot of the interaction effect indicates that as the product's utilitarian nature increased, basket value decreased, but the decrease was lower at websites with broad scope. Website scope mitigated the negative effect of the utilitarian characteristic.

functionality and basket value strengthens. Because utilitarian purchases involve reasoning, access to information and content is more positively associated with basket value.

Robustness Analyses

To enhance the validity of our results and improve confidence in our estimation approach, we estimated several alternative models (in addition to the specifications provided in Table 5, wherein the estimation approach is the same for various alternate models). Our proposed approach models a staged browsing and purchasing process, using a Type II Tobit model in which the selection stage is purchase incidence, conditional on which we observe basket value. Furthermore, page views and visit duration are modeled as endogenous regressors, and the four-equation system is jointly estimated with correlated errors and random effects specified at the customer level.

First, we alter this specification with a two-stage Heckman regression approach in which we maximize the joint likelihood (e.g., Bucklin and Sismeiro 2003) of the selection and the basket value stage. We treated page views and duration as endogenous, adopting the control function approach suggested by Petrin and Train (2010) and implemented in recent work in marketing (Fang, Lee, and Yang 2015). Thus, we first predict page views and visit duration in separate regressions and then use the errors as additional regressors in the basket value equation. Next, we use a two-step approach in which we first estimate the Mill's ratio using an instrumental variable probit and use it as a regressor in the basket value equation (e.g., Ying et al. 2006) in a three-equation system consisting of page views, duration, and basket value.

We also estimated an instrumental variable Type I Tobit model for the basket value equation in which the selection equation is not modeled explicitly. Then, we estimated a linear instrumental variable regression model estimated with two-stage least squares for the basket value equation and finally a crossed random effects model (that incorporates consumer as well as website-level random effects) for the basket value equation using a subset of 750 consumers from the sample (for details from these analyses, see the Web Appendix, Tables WA-4–WA-8). We also conducted in-sample and out-of-sample validation analysis, the results of which are presented in the Web Appendix (Table WA-10). Overall, we find that results pertaining to the effects of the primary variables (product and website heterogeneity) and their interactions are significant and consistent in magnitude and direction on the eventual outcome of interest, basket value, across various specifications. This result provides further validation for the robustness of our findings.

Discussion

When consumers shop online, they go through a staged browsing and purchasing experience in which a variety of factors influence various aspects of the purchase process. We propose that chief among these factors are the notions of what is being purchased (product) and where it is being purchased (website). These factors affect the browsing outcomes of page views and visit duration, which then influence the purchase decision, conditional on which a basket of certain value is realized. Recognizing the complexity of a typical online purchase experience, we develop an empirical model in which we estimate the independent and joint effects of product and website characteristics—or the what and where of the transaction—on the final outcome (i.e., basket value), after accounting for the staged nature of the process and controlling for consumer heterogeneity.

Implications

A key takeaway from our study is that the associations between product and website characteristics and online outcomes vary across outcomes. From a strategic standpoint, this finding is critical for online retailers: If their revenues do not depend on website advertising (i.e., are influenced by page views and visit duration) but rather on their e-commerce (i.e., basket value), retailers must be cognizant that improvements on one of the outcomes might hurt the other. Given the descriptive nature of our research, we are cautious in the interpretation of what this means to changing product assortments or modifying website layout and design. However, our findings provide some intriguing implications to relationships between factors viewed as crucial in the online shopping context.

Regarding the main effects of product characteristics on basket value, online retailers should pay greater attention to the degree of hedonic/utilitarian characteristics of the products they carry; we find that, all else being equal, a unit (standard deviation) increase in these characteristics is associated with a decrease in basket value by 4.9% ($4.80) for hedonic purchases and 3.1% ($2.95) for utilitarian purchases. Our data indicate that moderate to low levels of hedonic or utilitarian products are more favorable to online baskets than elevated levels of hedonic and functional baskets, perhaps because consumers anticipate greater guilt for products with high hedonic characteristics (Khan and Dhar 2010) and higher costs of failure for those with high utilitarian characteristics (Chitturi et al. 2007), which may result in lower basket value in an online environment (in which consumers judge these characteristics in a virtual setting).

Our findings regarding the main effects of website characteristics provide some significant insights regarding the role of the shopping platform itself. A website's product scope is positively related to purchase probability, visit duration, and basket value and negatively associated with page views. All else being equal, a broad-scope retailer's mean basket value is 2.9% ($2.82) higher in value than a narrow-scope retailer's. As online retailers continue to compete on razor-thin margins (Rabinovich, Rungtusanatham, and Laseter 2008), our findings reflect the broader trend in e-commerce of the rise of large dominant online retailers such as Amazon.com, which tend to offer consumers a one-stop shopping experience (Srinivasan and Moorman 2005).

We find some intriguing results related to website functionality. We find a negative association between communication functionality and purchase probability, page views, and basket value but positive association with visit duration. All else being equal, a unit (1 standard deviation) increase in a website's communication functionality is associated with a decrease of 3.4% ($3.28) in basket value. One possible explanation is that due to the pervasive use of online websites, consumers find certain communicative elements to be clutter (e.g., e-mail, chat rooms) and therefore buy less and view fewer pages, but inadvertently spend more time gathering information in areas such as the forums and Q&A sections. In contrast, navigational functionality is positively associated with purchase probability, basket value, and visit duration but negatively associated with page views. All else being equal, a unit increase in navigational functionality is associated with an increase of 6.6% ($6.35) in basket value. The ability to explore and and collect information without interacting with others seems to increase all outcomes except page views, perhaps because websites are creating new dynamic browsing elements such as videos and other visualization technologies (e.g., Hong et al. 2004) that do not require navigating to other pages. This raises some challenging questions as to what page views represent as a measure of consumer interest in the online environment. Our findings provide some preliminary evidence linking functionality to outcomes that suggest the need to develop more fine-grained measures of consumer interest.

The effects of product characteristics on basket value depend on the characteristics of the website. We summarize the direction of these interactive effects in Figure 3. A product's hedonic/utilitarian nature influences the extent to which the website's scope, in terms of product variety, can be associated with basket value: an average transaction at a broad-scope retailer had a basket value 38% ($16.20) higher than that at a narrow-scope retailer for products with high hedonic characteristics, such as jewelry and watches. Similarly, for purchases involving products with high utilitarian nature, such as office supplies, the difference in basket value was 36% ($19.25). For purchases of low hedonic and utilitarian products, the differences were 12% ($10.97) and 7.7% ($5.78), respectively. The wider assortment of categories carried by retailers whose websites have broad scope (González-Benito and Martos-Partal 2012) seemingly keeps consumers more engaged, such that they spend more money per transaction for both hedonic and utilitarian products. Our findings highlight the strategic role of product variety in influencing basket value and its increasing importance to online retailers as large retailers such as Wal-Mart expand their online presence.

The functionality of websites interacted with product characteristics to provide some insightful and intriguing results. We find that websites rich on communication functionality realize higher basket value for hedonic purchases. This finding supports our argument that when consumers are promotion focused in their purchase (Chitturi et al. 2007), seeking affirmation and support from others, the website's ability to encourage this improves basket value. However, navigational functionality helps realize higher basket value for utilitarian purchases but hurts hedonic purchases. This result is consistent with our view that certain aspects of the website that create clutter (Wells, Valacich, and Hess 2011) and do not add to the consumer's shopping experience might hurt the basket value. As consumers seek out different types of information for hedonic versus utilitarian purchases, aspects of the websites that facilitate this process rather than impede it affect shopping outcomes differently (Huang, Lurie, and Mitra 2009).

FIGURE 3
Summary of the Effects of Product and Website Characteristics on Basket Value

WC / PC	Website Scope	Communication Functionality	Navigational Functionality
Hedonic (–)[a]	↑	↑	↓
Utilitarian (–)[a]	↑	—	↑

[a]Indicates the direction of main effect of website and product characteristic variables on basket value. For example, communication functionality has a negative main effect on basket value.

Notes: WC = website characteristics; PC = product characteristics. The up arrows indicate a statistically significant, positive interaction effect. For example, in the first row, third column, the up arrow indicates that as the hedonic nature of products increases, it weakens the negative main effect of communication functionality on basket value. The down arrows indicate a statistically significant, negative interaction effect. For example, in the first row, fourth column, the down arrow indicates that as the hedonic nature of products increases, it weakens the positive effect of navigational functionality on basket value.

Contributions

To the best of our knowledge, the current research is the first large-scale empirical study to jointly explore the two stages in the online shopping context—browsing and purchasing. With our broad integrative approach, we contribute to an emerging stream that highlights the importance of large scale data analysis (Goel and Goldstein 2013; Kushwaha and Shankar 2013) and show that significant insights can be obtained by assembling richer data, as exemplified by our sample and additional data collected from a survey to calibrate the product category characteristics and multirater coding of the website characteristics (Danaher et al. 2006).

As online shopping becomes a dominant alternative to traditional shopping, the importance of the website's features is emerging as a critical influence on consumer trust and experience (Bart et al. 2005; Galletta et al. 2006). We contribute to this stream of work by offering a nuanced investigation of the two aspects of a website's functionality, communication and navigation, and enunciating their role in the different outcomes of the browsing and purchase process.

We also contribute to research on online shopping by exploring the factors that affect the two stages of the online shopping process, browsing and transaction (or purchase). In doing so, we explore interdependencies involved among various outcomes. We also explicitly consider the impact of product characteristics (hedonic and utilitarian) and website characteristics (scope and functionality), along with their joint influence. Thus, we capture variations in consumer behavior due to two critical determinants of behavior: the what and where of shopping. Although these aspects have received vast attention in empirical studies of traditional brick-and-mortar retailers (e.g., González-Benito and Martos-Partal 2012; Kushwaha and Shankar 2013), they have not been investigated sufficiently in online shopping contexts. In summary, we believe that with a hybrid data analytic approach such as ours, combining transaction-level secondary data with primary data collection approaches could provide researchers with more effective tools for understanding complex marketing issues.

Limitations and Directions for Research

We investigated online consumer behavior using session-level data. Although this allowed us to accommodate several factors that drive heterogeneity, it still requires further investigation. For example, because we do not explicitly observe which web pages were viewed during session not involving purchases, we could not infer the causal chain of events between browsing and purchasing (e.g., Huang, Lurie, and Mitra 2009), which raises the issue of last-click attribution. Future research could use our insights in developing clickstream models that incorporate such aspects of the browsing process. We incorporated many consumer-level characteristics, including demographics and prior online behavior, but other unobserved or omitted variables could be correlated with our explanatory variables, leading to potential bias. Although we mitigated this possibility by showing that the results are robust to alternative specifications that use instrumental variable approaches, we acknowledge the cross-sectional nature of our inquiry and do not make any causal claims regarding the proposed relationships.

Data limitations also did not allow us to address two important considerations: online marketing and the role of brands. Online retailers vary on how they attract and keep consumers loyal, and we could not account for variation across retailers on this dimension. Similarly, we also do not incorporate brand-level information or the possibility that seller ratings on websites might play a role in influencing consumers. We also do not observe consumers' offline behavior, a factor that might inform online search and purchase (Kushwaha and Shankar 2013). We suggest these limitations as opportunities for future research, as they offer some interesting challenges in terms of data compilation and model estimation. Finally, it would be worthwhile to merge online data with data on online feedback systems (e.g., Kumar et al. 2013) to explicate why consumers behave as they do in online buying contexts.

As online commerce evolves, it is important for firm that adopt this medium to understand the role of the contextual factors in influencing consumer behavior (Grewal et al. 2009). We investigate this important issue using a unique data set and develop a model to capture the whole range of issues involved in the phenomenon. As a result, we unearth implications for academic research in the domain and importantly, insights for managerial practice. We hope our research sets the trend for other comprehensive studies in this domain.

REFERENCES

Bart, Yakov, Venkatesh Shankar, Fareena Sultan, and Glen L. Urban (2005), "Are the Drivers and Role of Online Trust the Same for All Web Sites and Consumers? A Large-Scale Exploratory Empirical Study," *Journal of Marketing*, 69 (October), 133–52.

Bell, David R. and James M. Lattin (1998), "Shopping Behavior and Consumer Preference for Store Price Format: Why "Large Basket" Shoppers Prefer EDLP," *Marketing Science*, 17 (1), 66–88.

Benartzi, Shlomo, Roni Michaely, and Richard Thaler (1997), "Do Changes in Dividends Signal the Future or the Past?" *Journal of Finance*, 52 (3), 1007–34.

Bhargave, Rajesh and Nicole Votolato Montgomery (2013), "The Social Context of Temporal Sequences: Why First Impressions Shape Shared Experiences," *Journal of Consumer Research*, 40 (3), 501–17.

Boatwright, Peter, Sharad Borle, and Joseph B. Kadane (2003), "A Model of the Joint Distribution of Purchase Quantity and Timing," *Journal of the American Statistical Association*, 98 (463), 564–72.

Brandt, Richard L (2011), *One Click: Jeff Bezos and the Rise of Amazon.com*. New York: Penguin.

Bucklin, Randolph E. and Catarina Sismeiro (2003), "A Model of Web Site Browsing Behavior Estimated on Clickstream Data," *Journal of Marketing Research*, 40 (August), 249–67.

Chandukala, Sandeep R., Jeffrey P. Dotson, Jeff D. Brazell, and Greg M. Allenby (2011), "Bayesian Analysis of Hierarchical Effects," *Marketing Science*, 30 (1), 123–33.

Chernev, Alexander (2004), "Goal-Attribute Compatibility in Consumer Choice," *Journal of Consumer Psychology*, 14 (1), 141–50.

Chitturi, Ravindra, Rajagopal Raghunathan, and Vijay Mahajan (2007), "Form Versus Function: How the Intensities of Specific Emotions Evoked in Functional Versus Hedonic Trade-Offs Mediate Product Preferences," *Journal of Marketing Research*, 44 (November), 702–14.

———, ———, and ——— (2008), "Delight by Design: The Role of Hedonic Versus Utilitarian Benefits," *Journal of Marketing*, 72 (May), 48–63.

ComScore (2011), "Web Behavior Disaggregated: Dataset," University of Pennsylvania Wharton Research Data Services.

Cragg, John G. and Stephen G. Donald (1993), "Testing Identifiability and Specification in Instrumental Variable Models," *Econometric Theory*, 9 (02), 222–40.

Cuddihy, Elisabeth and Jan H. Spyridakis (2012), "The Effect of Visual Design and Placement of Intra-article Navigation Schemes on Reading Comprehension and Website User Perceptions," *Computers in Human Behavior*, 28 (4), 1399–409.

Danaher, Peter J. (2007), "Modeling Page Views Across Multiple Websites with an Application to Internet Reach and Frequency Prediction," *Marketing Science*, 26 (3), 422–37.

———, Guy W. Mullarkey, and Skander Essegaier (2006), "Factors Affecting Web Site Visit Duration: A Cross-Domain Analysis," *Journal of Marketing Research*, 43 (May), 182–94.

Drukker, David M. and Richard Gates (2006), "Generating Halton Sequences Using Mata," *Stata Journal*, 6 (2), 214–28.

Fader, Peter S. and Bruce G.S. Hardie (1996), "Modeling Consumer Choice Among SKUs," *Journal of Marketing Research*, 33 (November), 442–52.

Fang, Eric, Jongkuk Lee, and Zhi Yang (2015), "The Timing of Codevelopment Alliances in New Product Development Processes: Returns for Upstream and Downstream Partners," *Journal of Marketing*, 79 (January), 64–82.

Galletta, Dennis F., Raymond M. Henry, Scott McCoy, and Peter Polak (2006), "When the Wait Isn't So Bad: The Interacting Effects of Website Delay, Familiarity, and Breadth," *Information Systems Research*, 17 (1), 20–37.

Garg, Nitika, J. Jeffrey Inman, and Vikas Mittal (2005), "Incidental and Task-related Affect: A Re-inquiry and Extension of the Influence of Affect on Choice," *Journal of Consumer Research*, 32 (1), 154–59.

Ghose, Sanjoy and Wenyu Dou (1998), “Interactive Functions and Their Impacts on the Appeal of Internet Presence Sites,” *Journal of Advertising Research*, 38 (2), 29–44.
Goel, Sharad and Daniel G. Goldstein (2013), “Predicting Individual Behavior with Social Networks,” *Marketing Science*, 33 (1), 82–93.
Goldsmith, Kelly, Eunice Kim Cho, and Ravi Dhar (2012), “When Guilt Begets Pleasure: The Positive Effect of a Negative Emotion,” *Journal of Marketing Research*, 49 (December), 872–81.
González-Benito, Óscar and Mercedes Martos-Partal (2012), “Role of Retailer Positioning and Product Category on the Relationship between Store Brand Consumption and Store Loyalty,” *Journal of Retailing*, 88 (2), 236–49.
Grewal, Dhruv, Michael Levy, and V. Kumar (2009), “Customer Experience Management in Retailing: An Organizing Framework,” *Journal of Retailing*, 85 (1), 1–14.
Higgins, E. Tory (1997), “Beyond Pleasure and Pain,” *American Psychologist*, 52 (12), 1280–1300.
Hong, Weiyin, James Y.L. Thong, and Kar Yan Tam (2004), “Does Animation Attract Online Users’ Attention? The Effects of Flash on Information Search Performance and Perceptions,” *Information Systems Research*, 15 (1), 60–86.
Huang, Peng, Nicholas H. Lurie, and Sabyasachi Mitra (2009), “Searching for Experience on the Web: An Empirical Examination of Consumer Behavior for Search and Experience Goods,” *Journal of Marketing*, 73 (March), 55–69.
Inman, J. Jeffrey, Russell S. Winer, and Rosellina Ferraro (2009), “The Interplay Among Category Characteristics, Customer Characteristics, and Customer Activities on In-store Decision Making,” *Journal of Marketing*, 73 (September), 19–29.
Johnson, Eric J., Wendy W. Moe, Peter S. Fader, Steven Bellman, and Gerald L. Lohse (2004), “On the Depth and Dynamics of Online Search Behavior,” *Management Science*, 50 (3), 299–308.
Khan, Uzma and Ravi Dhar (2006), “Licensing Effect in Consumer Choice,” *Journal of Marketing Research*, 43 (May), 259–66.
——— and ——— (2010), “Price-Framing Effects on the Purchase of Hedonic and Utilitarian Bundles,” *Journal of Marketing Research*, 47 (December), 1090–99.
Krasnikov, Alexander and Satish Jayachandran (2008), “The Relative Impact of Marketing, Research-and-Development, and Operations Capabilities on Firm Performance,” *Journal of Marketing*, 72 (July), 1–11.
Kumar, V., Vikram Bhaskaran, Rohan Mirchandani, and Milap Shah (2013), “Creating a Measurable Social Media Marketing Strategy: Increasing the Value and ROI of Intangibles and Tangibles for Hokey Pokey,” *Marketing Science*, 32 (2), 194–212.
Kushwaha, Tarun and Venkatesh Shankar (2013), “Are Multichannel Customers Really More Valuable? The Moderating Role of Product Category Characteristics,” *Journal of Marketing*, 77 (July), 67–85.
Landwehr, Jan R, Daniel Wentzel, and Andreas Herrmann (2013), “Product Design for the Long Run: Consumer Responses to Typical and Atypical Designs at Different Stages of Exposure,” *Journal of Marketing*, 77 (September), 92–107.
Li, Hongshuang and P.K. Kannan (2014), “Attributing Conversions in a Multichannel Online Marketing Environment: An Empirical Model and a Field Experiment,” *Journal of Marketing Research*, 51 (February), 40–56.
Lin, Lin, Paul Jen-Hwa Hu, Olivia R. Liu Sheng, and Johnny Lee (2010), “Is Stickiness Profitable for Electronic Retailers?” *Communications of the ACM*, 53 (3), 132–36.
Lindstrom, Mary J. and Douglas M. Bates (1990), “Nonlinear Mixed Effects Models for Repeated Measures Data,” *Biometrics*, 46 (3), 673–87.
Moe, Wendy W. (2006), “An Empirical Two-Stage Choice Model with Varying Decision Rules Applied to Internet Clickstream Data,” *Journal of Marketing Research*, 43 (November), 680–92.
Montgomery, Alan L., Shibo Li, Kannan Srinivasan, and John C. Liechty (2004), “Modeling Online Browsing and Path Analysis Using Clickstream Data,” *Marketing Science*, 23 (4), 579–95.
Narayanan, Sridhar and Kirthi Kalyanam (2015), “Position Effects in Search Advertising and their Moderators: A Regression Discontinuity Approach,” *Marketing Science*, 34 (3), 388–407.
Okada, Erica Mina (2005), “Justification Effects on Consumer Choice of Hedonic and Utilitarian Goods,” *Journal of Marketing Research*, 42 (February), 43–53.
O’Neill, S. and J. Chu (2001), “Modeling to Scale: The Evolution of Online Grocery Retailing,” IBM Global Services eStrategy Report.
Petersen, J. Andrew and V. Kumar (2009), “Are Product Returns a Necessary Evil? Antecedents and Consequences,” *Journal of Marketing*, 73 (May), 35–51.
Petrin, Amil and Kenneth Train (2010), “A Control Function Approach to Endogeneity in Consumer Choice Models,” *Journal of Marketing Research*, 47 (February), 3–13.
Rabinovich, Elliot, Manus Rungtusanatham, and Timothy M. Laseter (2008), “Physical Distribution Service Performance and Internet Retailer Margins: The Drop-Shipping Context,” *Journal of Operations Management*, 26 (6), 767–80.
Raghu, T.S., P.K. Kannan, H. Raghav Rao, and Andrew B. Whinston (2001), “Dynamic Profiling of Consumers for Customized Offerings Over the Internet: A Model and Analysis,” *Decision Support Systems*, 32 (2), 117–34.
Richins, Marsha L (2013), “When Wanting Is Better than Having: Materialism, Transformation Expectations, and Product-Evoked Emotions in the Purchase Process,” *Journal of Consumer Research*, 40 (1), 1–18.
Roodman, Davis (2011), “Estimating Fully Observed Recursive Mixed-Process Models with CMP,” *Stata Journal*, 11 (2), 159–206.
Rose, Susan, Moira Clark, Phillip Samouel, and Neil Hair (2012), “Online Customer Experience in e-Retailing: An Empirical Model of Antecedents and Outcomes,” *Journal of Retailing*, 88 (2), 308–22.
Shankar, Venkatesh, Amy K. Smith, and Arvind Rangaswamy (2003), “Customer Satisfaction and Loyalty in Online and Offline Environments,” *International Journal of Research in Marketing*, 20 (2), 153–75.
Sismeiro, Catarina and Randolph E. Bucklin (2004), “Modeling Purchase Behavior at an e-Commerce Web Site: A Task-Completion Approach,” *Journal of Marketing Research*, 41 (August), 306–23.
Sridhar, Shrihari and Raji Srinivasan (2012), “Social Influence Effects in Online Product Ratings,” *Journal of Marketing*, 76 (September), 70–88.
Srinivasan, Raji and Christine Moorman (2005), “Strategic Firm Commitments and Rewards for Customer Relationship Management in Online Retailing,” *Journal of Marketing*, 69 (October), 193–200.
Swaminathan, Anand (1995), “The Proliferation of Specialist Organizations in the American Wine Industry, 1941–1990,” *Administrative Science Quarterly*, 40 (4), 653–80.
Townsend, Claudia and Sanjay Sood (2012), “Self-Affirmation Through the Choice of Highly Aesthetic Products,” *Journal of Consumer Research*, 39 (2), 415–28.

Verbeke, Geert and Emmanuel Lesaffre (1996), "A Linear Mixed-Effects Model with Heterogeneity in the Random-Effects Population," *Journal of the American Statistical Association*, 91 (433), 217–21.

Voss, Kevin E., Eric R. Spangenberg, and Bianca Grohmann (2003), "Measuring the Hedonic and Utilitarian Dimensions of Consumer Attitude," *Journal of Marketing Research*, 40 (August), 310–20.

Wells, John D., Joseph S. Valacich, and Traci J. Hess (2011), "What Signals Are You Sending? How Website Quality Influences Perceptions of Product Quality and Purchase Intentions," *Management Information Systems Quarterly*, 35 (2), 373–96.

Yadav, Manjit S. and Paul A. Pavlou (2014), "Marketing in Computer-Mediated Environments: Research Synthesis and New Directions," *Journal of Marketing*, 78 (January), 20–40.

Ying, Yuanping, Fred Feinberg, and Michel Wedel (2006), "Leveraging Missing Ratings to Improve Online Recommendation Systems," *Journal of Marketing Research*, 43 (August), 355–65.

Young, Martin R., Wayne S. DeSarbo, and Vicki G. Morwitz (1998), "The Stochastic Modeling of Purchase Intentions and Behavior," *Management Science*, 44 (2), 188–202.

Zhang, Xiaoling, Shibo Li, Raymond R. Burke, and Alex Leykin (2014), "An Examination of Social Influence on Shopper Behavior Using Video Tracking Data," *Journal of Marketing*, 78 (September), 24–41.

Zhu, Feng and Xiaoquan Zhang (2010), "Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics," *Journal of Marketing*, 74 (March), 133–48.

WEB APPENDIX

Exploring the Effects of What (Product) and Where (Website) Characteristics on Online Shopping Behavior

Girish Mallapragada
Sandeep R. Chandukala
Qing Liu

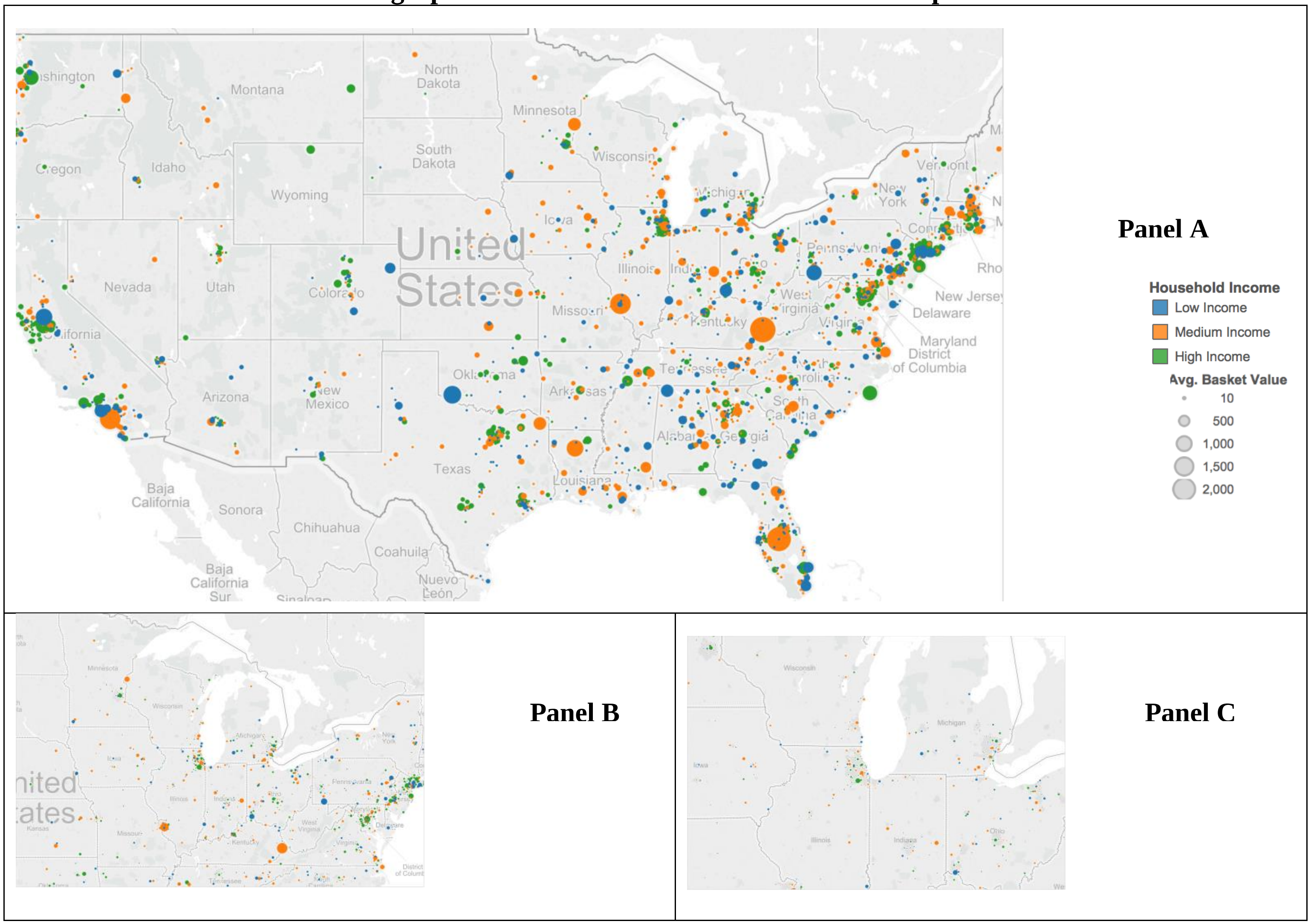
Visualization of Demographics and Key Variables

Visualization of the data is one mechanism by which we build a more detailed understanding of the characteristics of the data (Lilien, Roberts, and Shankar 2013) and by using multiple visualization schemes we show that the sample is well represented, giving us confidence in making inferences about the general population of online shoppers.

To demonstrate the volume and geographical distribution of our browsing data, in Figure A. 1, we created three maps at different detail levels that show the geographic distribution of the online transactions in our sample. Panel A shows the detail for the whole of United States, Panel B shows the detail for Mid-West and Panel C shows the detail for the Great Lakes region. The location of the circles represents the zip codes in which the online transactions occurred, their size represents the average basket value per transaction and their color represents the income category of the household from which the transactions occurred. We aim to develop a nuanced understanding of consumer behavior in the online context because there is a critical need to start applying a big data analytic approach by not restricting the sample or the scope of the phenomenon. By doing so, we cover as much ground as possible in trying to model the complex reality of online shopping.

In Figure A.2, we created stacked bar charts depicting the average page views, duration of visit and basket value broken down across the various categories of the three demographic characteristics - household size, household income and racial background.

Figure WA-1
Geographic Distribution of Households in the Sample



Notes: 1) Panel A shows the distribution detail for US, Panel B for mid-west and Panel C for parts of Great Lakes region

Figure WA-2
Summary Characteristics of Page Views, Duration of Visit and Basket Value

Notes: 1) The stacked bars, each color representing a race, show the average page views, duration of visit and basket value across the three income groups cross-tabulated against the three household sizes.

Table WA-1
Comparison with Relevant Research on modeling online browsing and purchasing behavior

References	Dependent variables	Covariates	Model features
Our Research	Basket value Purchase incidence Page views Duration	Website characteristics Consumer characteristics Product characteristics Past browsing Past purchase	Endogeneity accounted system of equations with Correlated errors
Danaher, Mullarkey and Essegaier (2006)	Page views Duration	Website characteristics Consumer characteristics Visit occasions	Separate equations
Bart, Shanker, Sultan And Urban (2005)	Trust Purchase intent	Website characteristics Consumer characteristics	Separate equations
Bucklin and Sismeiro (2003)	Page request Duration	Website characteristics Past browsing	Correlated errors
Sismeiro and Bucklin (2004)	Purchase incidence (as a series of task completion)	Page views Duration Website characteristics Visit occasions	Conditional models
Moe and Fader (2004)	Purchase probability	Past browsing Past purchase	Dynamic probability model
Montgomery, Li, Srinivasan and Liechty (2004)	Browsing path	Past browsing	Dynamic MNP model
Moe (2006)	Viewing choice Purchase decision	Product attributes	Two-stage choice model

Table WA-2
Distribution of Households Across Key Demographics

		Household Size			
Income	**Race**	**Small**	**Medium**	**Large**	**Totals (%) by Income**
Low	**Caucasian**	172	136	76	**579 (29%)**
	African-American	55	71	54	
	Asian	5	7	3	
Medium	**Caucasian**	240	215	130	**780 (39%)**
	African-American	45	57	69	
	Asian	6	7	11	
High	**Caucasian**	200	220	101	**641 (32%)**
	African-American	32	31	19	
	Asian	20	14	4	
Totals (%) by Household Size		**775 (39%)**	**758 (38%)**	**467 (23%)**	**2000**
Totals (%) by Race				**Caucasian**	**1490 (75%)**
				African-American	**433 (22%)**
				Asian	**77 (4%)**
				Total Number of Households	**2000**

Notes: This table breaks down the unique households in the sample by income category (rows) and, within each income category, by the racial background of the head of household. The income category level totals are in the final column labeled "Totals (%) by Income." The breakdown by household size within each income–race sub-category appears in the columns. The household size totals are in the penultimate row, labeled "Totals (%) by Household Size." Finally, the breakdown by racial background appears in the final row, with total labeled "Totals (%) by Race."

Table WA-3
Effects of customer characteristics in the conditional purchase model

Customer Characteristic	Basket Value
Medium Household Size	0.062*** (0.027)
Large Household Size	0.181*** (0.029)
Medium Household Income	-0.017 (0.026)
Large Household Income	0.001 (0.032)
African-American	0.149*** (0.031)
Asian	0.358*** (0.068)
Northcentral	-0.269*** (0.037)
South	-0.129*** (0.033)
West	-0.186 (0.041)
Connection Speed	0.336*** (0.079)

Table WA-4
Estimates from Heckman Regression

		Probability of Purchase	Basket Value
Intercept		-3.177*** (0.142)	4.006*** (0.136)
Browsing Behavior	**Page Views**	0.095** (0.049)	0.239*** (0.059)
	Duration	0.274*** (0.036)	-0.297*** (0.056)
Product Characteristics	**Hedonic**	--	-0.279*** (0.024)
	Utilitarian	--	-0.142*** (0.021)
Website Characteristics	**Website Scope**	0.097*** (0.018)	0.181*** (0.029)
	Comm. Functionality	-0.314*** (0.013)	-0.184*** (0.071)
	Nav. Functionality	0.002 (0.010)	0.312*** (0.032)
Interaction between Product and Website Characteristics	**WS X Hedonic**	--	0.192***(0.034)
	WS X Utilitarian	--	0.125*** (0.025)
	CF X Hedonic	--	0.163*** (0.057)
	CF X Utilitarian	--	-0.035 (0.032)
	NF X Hedonic	--	-0.153*** (0.023)
	NF X Utilitarian	--	0.134*** (0.055)
Control Variables	**Price Dispersion**	--	0.000*** (0.000)
	Basket Variety	--	0.179*** (0.010)
	Domain Variety	-0.002*** (0.000)	-0.001*** (0.000)
	Past Incidence	0.121*** (0.002)	--
	Past Purchases	--	0.018*** (0.003)
	Customer Characteristics	Included	Included
	Time Dummies	Included	Included
Mill's Ratio		**--**	-0.118*** (0.022)

Notes: 1) Bootstrap standard errors are reported as part of two stage estimation involving control function residuals
2) Customer Characteristics include dummies for household size, income, racial background, census region and internet connection type.
3) * : $p < .10$, ** : $p < .05$, *** : $p < .01$

Table WA-5
Estimates from Multivariate Mixed-effects Two-step Regression

		Page Views	Duration	Basket Value
Intercept		1.301*** (0.017)	1.290*** (0.021)	3.795*** (0.115)
Browsing Behavior	**Page Views**	--	--	0.174*** (0.018)
	Duration	--	--	-0.127***(0.017)
Product Characteristics	**Hedonic**	--	--	-0.272*** (0.026)
	Utilitarian	--	--	-0.140*** (0.020)
Website Characteristics	**Website Scope**	-0.160*** (0.003)	0.080*** (0.005)	0.150*** (0.027)
	Comm. Functionality (CF)	-0.031***(0.001)	0.041*** (0.001)	-0.187*** (0.064)
	Nav. Functionality (NF)	-0.040*** (0.005)	0.049*** (0.005)	0.297*** (0.026)
Interaction between Product and Website Characteristics	**WS X Hedonic**	--	--	0.181***(0.033)
	WS X Utilitarian	--	--	0.121*** (0.029)
	CF X Hedonic	--	--	0.164*** (0.051)
	CF X Utilitarian	--	--	-0.035 (0.030)
	NF X Hedonic	--	--	-0.151*** (0.021)
	NF X Utilitarian	--	--	0.134*** (0.054)
Control Variables	**Price Dispersion**	--	--	0.000*** (0.000)
	Tot. Basket Items	--	--	0.004*** (0.003)
	Basket Variety	--	--	0.173*** (0.010)
	Domain Variety	0.000*** (0.000)	0.000*** (0.000)	-0.002*** (0.000)
	Past Page Views	0.026*** (0.001)	--	--
	Past Duration	--	0.022*** (0.000)	--
	Past Purchases	--	--	0.017*** (0.002)
	Customer Characteristics	Included	Included	Included
	Time Dummies	Included	Included	Included
Mill's Ratio			--	-0.124*** (0.025)

Notes: 1) Mill's Ratio was obtained using an instrumental variable probit regression
2) Customer Characteristics include dummies for household size, income, racial background, census region and internet connection type.
3) * : p < .10, ** : p < .05, *** : p < .01

Table WA-6
Estimates from Instrumental Variable Tobit Regression

		Basket Value
Intercept		4.138*** (0.292)
Browsing Behavior	**Page Views**	0.364*** (0.082)
	Duration	-0.470***(0.106)
Product Characteristics	**Hedonic**	-0.307*** (0.041)
	Utilitarian	-0.131*** (0.026)
Website Characteristics	**Website Scope (WS)**	0.128*** (0.037)
	Comm. Functionality (CF)	-0.231*** (0.043)
	Nav. Functionality (NF)	0.301*** (0.026)
Interaction between Product and Website Characteristics	**WS X Hedonic**	0.207***(0.047)
	WS X Utilitarian	0.125*** (0.031)
	CF X Hedonic	0.174*** (0.042)
	CF X Utilitarian	-0.035 (0.026)
	NF X Hedonic	-0.160*** (0.021)
	NF X Utilitarian	0.156*** (0.029)
Control Variables	**Price Dispersion**	0.000*** (0.000)
	Basket Variety	0.225*** (0.018)
	Domain Variety	-0.002*** (0.000)
	Past Purchases	0.017*** (0.001)
	Customer Characteristics	Included
	Time Dummies	Included

Notes: 1) Page Views and Duration were instrumented using past rolling average, website and customer characteristics, and time dummies.
2) Customer Characteristics include dummies for household size, income, racial background, census region and internet connection type.
3) * : $p < .10$, ** : $p < .05$, *** : $p < .01$

Table WA-7
Estimates from Instrumental Variable Linear Regression

		Basket Value
Intercept		4.348*** (0.407)
Browsing Behavior	**Page Views**	0.126*** (0.059)
	Duration	-0.167*** (0.061)
Product Characteristics	**Hedonic**	-0.296*** (0.033)
	Utilitarian	-0.142*** (0.020)
Website Characteristics	**Website Scope**	0.154*** (0.031)
	Comm. Functionality (CF)	-0.141***(0.051)
	Nav. Functionality (NF)	0.305*** (0.023)
Interaction between Product and Website Characteristics	**WS X Hedonic**	0.198*** (0.038)
	WS X Utilitarian	0.125*** (0.027)
	CF X Hedonic	0.157*** (0.037)
	CF X Utilitarian	-0.033* (0.024)
	NF X Hedonic	-0.159*** (0.019)
	NF X Utilitarian	0.143*** (0.025)
Control Variables	**Price Dispersion**	0.000*** (0.000)
	Basket Variety	0.017*** (0.001)
	Domain Variety	-0.002*** (0.000)
	Past Purchases	0.016*** (0.001)
	Customer Characteristics	Included
	Time Dummies	Included
Mill's Ratio		-0.274*** (0.097)

Notes: 1) Estimates obtained using 2SLS
2) Mill's ratio was estimated using an instrumental variable probit regression
3) 3) * : p < .10, ** : p < .05, *** : p < .01

Table WA-8
Estimates from Crossed Random Effects Regression

		Basket Value
Intercept		3.616*** (0.486)
Browsing Behavior	**Page Views**	0.162*** (0.052)
	Duration	-0.070*** (0.046)
Product Characteristics	**Hedonic**	-0.261*** (0.033)
	Utilitarian	-0.249*** (0.020)
Website Characteristics	**Website Scope**	0.140 (0.180)
	Comm. Functionality (CF)	-0.593***(0.051)
	Nav. Functionality (NF)	0.385*** (0.023)
Interaction between Product and Website Characteristics	**WS X Hedonic**	0.199*** (0.100)
	WS X Utilitarian	0.081*** (0.063)
	CF X Hedonic	0.378*** (0.116)
	CF X Utilitarian	-0.180* (0.051)
	NF X Hedonic	-0.206* (0.147)
	NF X Utilitarian	0.504*** (0.139)
Control Variables	**Price Dispersion**	0.000*** (0.000)
	Basket Variety	0.197*** (0.028)
	Domain Variety	-0.002*** (0.001)
	Past Purchases	0.018*** (0.002)
	Customer Characteristics	Included
	Time Dummies	Included
Mill's Ratio		-0.033*** (0.070)
Random Effect (Domain)		0.467 (0.238)
Random Effect (Customer)		0.525 (0.247)

Notes: 1) Estimates obtained using maximum likelihood with Laplacian approximation for random effects
2) Mill's ratio was estimated using an instrumental variable probit regression
3) 3) * : p < .10, ** : p < .05, *** : p < .01

Table WA-9
Factor Loading Matrix for Website Functionality

Items (website features)	Communication Functionality	Website Functionality
Change language	0.309	**0.574**
Change Graphics or Text	0.026	**0.608**
Change Page Layout	-0.062	**0.817**
Customize Site Content	-0.052	**0.783**
E-mail Contact	**0.518**	-0.065
View Product/Service Information	0.458	**0.639**
Online Help	**0.633**	0.367
Basic Search	0.079	**0.556**
Site Map	0.018	**0.526**
Links to Other Areas on Website	0.055	**0.703**
Download Content	0.389	**0.545**
Registration	**0.534**	0.081
Feedback via online forms	**0.631**	0.14
Feedback via email	**0.537**	0.249
Online Diagnostics	0.117	**0.514**
Recent Updates	0.164	**0.702**
Chat Rooms	**0.695**	0.016
Topic-specific Discussion Forums	**0.612**	0.323
Message Boards	**0.678**	0.393

Notes: 1) Varimax rotated loadings are reported

Table WA-10
Validation Analysis

Fit-statistic	In-Sample (N1 = 2000)		Out of Sample (N2= 2000)	
	Benchmark Model	Proposed Model	Benchmark Model	Proposed Model
MSE	0.73	**0.66**	2.01	**1.05**
MAPE	0.18	**0.16**	0.24	**0.19**
R-square	0.26	**0.35**	0.08	**0.17**