ABSTRACT

Title of Document:	**UNDERSTANDING CONSUMERS' ONLINE INFORMATION RETRIEVAL AND SEARCH: IMPLICATIONS FOR FIRM STRATEGIES**
	Vandana Ramachandran **Doctor of Philosophy** **2010**
Directed By:	**Prof. Siva Viswanathan** Decisions, Operations and Information Technology **Prof. Hank Lucas** Decisions, Operations and Information Technology

The growth of the Internet and other digitization technologies has enabled the unbundling of the physical and information components of the value chain and has led to an explosion of information made available to consumers. Understanding the implications of this new informational landscape for theory and practice is one of the key objectives of my research. My dissertation seeks to understand how firms can use their knowledge of online consumer search and information seeking behaviors to design optimal information provision strategies. The main premise is that consumers' online search behaviors are key to understanding consumers' underlying information needs and preferences. In my first essay I specifically focus on big-ticket, high-involvement goods for which firms essentially have sparse

information on their potential buyers – making information reflected in consumers' online search very valuable to online retailers. I use a new and rich source of clickstream data obtained from a leading clicks-and-mortar retailer to model consumers' purchase outcomes as a function of the product and price information provided by the retailer, and find interesting differences for sessions belonging to customers classified as browsers, directed shoppers and deliberating researchers. Since consumers typically straddle online as well as traditional channels, the second essay in my dissertation examines how online information acquired by consumers affects their choices in offline used-good markets. Secondary markets characterized by information asymmetries have typically resorted to quality-signaling mechanisms such as certification to help reduce the associated frictions. However, the value of traditional quality signals to consumers depends crucially on the extent of the asymmetries in these markets. The online information available to consumers today may help bridge such asymmetries. Drawing upon a unique and extensive dataset of over 12,000 consumers who purchased used vehicles, I examine the impact of their information acquisition from online intermediaries on their choice of (reliance on) one such quality signal - certification, as well as the price paid. These findings will help firms to better understand how the provision of different types of online information impacts consumers' choices and outcomes, and therefore help them in designing better and targeted strategies to interact with consumers.

UNDERSTANDING CONSUMERS' ONLINE INFORMATION RETRIEVAL AND SEARCH: IMPLICATIONS FOR FIRM STRATEGIES

By

Vandana Ramachandran

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2010

Advisory Committee:
Prof. Siva Viswanathan, Chair
Prof. Hank Lucas
Prof. Ritu Agarwal
Prof. Jie Zhang
Prof. Ingmar Prucha

Dedication

I'd like to dedicate this work to my family –mom, dad, and sister – who have been there

for me in so many ways throughout my long journey to get to this point. Thank you for

your patience, encouragement and belief in me.

Acknowledgements

First and foremost, I'd like to express my deepest gratitude for the valuable guidance and advice that I received from Siva all along the road to completing my dissertation. I have been amazingly fortunate to have an advisor whose patience and support has helped me overcome and recover from many situations. I'd like to thank Hank for patiently reading through my drafts and offering his insights across several rounds. Next, I'd like to thank Ritu for the wonderful seminars that challenged me to think broad and reminded me to look at the big picture in research, especially when I found myself lost in little details. I am very grateful to Jie and Ingmar for serving as committee members, and for their technical insights that have proved valuable in helping me complete this work. I have truly enjoyed being part of the IS group at Maryland – I thank the faculty there for interesting seminars and discussions that have helped me think about my research. I also want to thank Arun for his support and advice during the final leg of my journey. Finally, I will cherish the fun times spent with my PhD colleagues at the Robert H Smith School of Business - the ADS lunches, the volleyball games, the soccer matches, the late night coffee runs, the all-niters and the wonderful discussions that we shared about research and life. All of you have inspired me in countless ways, and I only hope to be able to give back to the community what you have given me.

Table of Contents

Chapter 1: Introduction and Overview

A good understanding of consumer behavior is the cornerstone of a firm's strategy. Naturally, firms have invested significantly in information systems ranging from point-of-sale scanners to RFID tags to gather and analyze information relating to consumer shopping behavior and purchase patterns. While these information technologies have significantly increased our understanding of consumer behavior in traditional channels, their potential pales in comparison with the amount and granularity of consumer-related information available through online channels. Online retailers today have the ability to gather fine-grained information about consumer behavior that can help fine tune their strategies to target individual customers and micro-segments in unique ways.

My dissertation seeks to understand how firms can use their knowledge of consumers' online search behaviors to design optimal strategies for the provision of information to customers. The underlying premise is that the online search behaviors of consumers reflect their underlying *information needs* and *preferences*. Online retailers and market intermediaries can then leverage this knowledge to build customized interactions with consumers and ultimately, influence their purchase outcomes. Within this framework, I describe two essays that examine related questions.

It is common today for consumers to search online to learn about the assortment of available products, brands and prices across firms – information that is likely to significantly influence their purchase outcomes. A critical challenge for online retailers therefore is to determine what types of product and price information are best suited to influence customers to purchase. While it has long been known that not all customers are in the same state of shopping when they visit a retailer, it has been cumbersome or

impossible to learn about such latent or unobserved differences in traditional settings. The availability of clickstream data however can potentially solve this problem by allowing retailers to construct meaningful segments of customer sessions on the basis of behaviors observed at their website. The goal of this research is twofold. First, consumers' search and navigation behaviors gleaned from clickstream data are used to meaningfully characterize consumers in ways that reflect shopping-relevant underlying differences across sessions. These differences are referred to as *states of shopping*. Second, I examine whether product and price-related online information had different impacts on conversion for customers belonging to various *states of shopping*, and whether information varied in its impacts on purchase related behaviors *within a session* (complete the purchase) and *across sessions* (return visit and future purchase). A three-state model comprised of *directed shoppers, deliberating researchers* and *browsers*, best describes the latent differences across customers shopping for big-ticket durable goods at a large US retailer. This categorization allows us to then uncover important differences in the effects of three types of product and price information across the three types of sessions. An interesting aspect of this model is that by allowing customers to belong to different latent states of shopping across repeat sessions to the retailer, we are able to uncover tradeoffs or contrasting effects of information on within-session conversion versus across-session purchase-related behaviors. The results provide evidence that questions the current common practice of offering price promotions such as free shipping and product category discounts to all customers that are visiting a store, and highlights the ability of rich product information to increase the stickiness of the website and loyalty of its customers. Moreover, depending on the retailer's goal – immediate conversion in

the short term versus developing a longer-term relationship with its customers– it is shown show that a different information provision strategy is likely to be optimal.

In the second essay, given that consumers typically straddle online as well traditional channels, I examine the cross-channel impacts of obtaining different types of online information on consumers' purchase outcomes in a traditional market. Specifically, the role of information is examined in the market for used goods where consumers face ample information asymmetry; and where sellers have typically resorted to selling quality signals (for a premium) to help reduce the associated information frictions. In recent times however there has been a tremendous growth in the volume and type of information related to all aspects of purchasing used vehicles that is available in online channels. I specifically study how the increased access to and use of decentralized online information related to product and price alters the value of one such centrally provided source of information - certification – in the used car market. It is theorized that information has both a first order effect on the expected quality or value of the used vehicle, and a second order effect on the perceived differentiation between certified and non-certified used vehicles- which combine to produce varying effects. Using a unique dataset of consumers who obtained vehicle and transaction related information from online sources in their used vehicle purchase process, the impact of their information acquisition on the choice of vehicle (certified or not), as well as the price paid is examined. The preliminary findings from this essay highlights the nature of complements/substitutes that emerge among buyers' search for online information related to the purchase, their reliance on traditional quality signals, and the price paid. In particular, it is found that product-related information substitutes, while price-related

information complements, certification, as indicated by their differential impacts on the demand and price of certified and uncertified used cars. The relevance of these findings for buyers and sellers are discussed and implications for online information providers are outlined as well.

Together the findings from these studies will allow us to gain a deeper understanding of how consumers' online information search behaviors are related to outcomes in online as well as traditional markets, and their implications for firms' strategies for information provision.

Chapter 2: State of Shopping and the Value of Information: Insights from the Clickstream

2.1. Introduction and motivation

Provision of targeted information is ubiquitous on the Internet today, and exists in myriad forms across search engines, online social networks, blogs, and various content sites. Much anecdotal evidence points to the positive effects of targeting done correctly-satisfied users and improved conversion rates. In this study, we extend the notion of targeting to *product and price-related information* that retailers can present in *real time* to customers who are *actively visiting* their online store. This micro-level approach has the potential to be highly interactive and complementary to targeting strategies used to attract consumers to e-tailer stores. Real-time targeting requires retailers to present custom information that matches their customers' needs and preferences, which are in turn driven by their shopping state during a session. Consumer's state of shopping is however unobserved, requiring retailers to make inferences on the basis of observed behavior patterns of consumers. A commonly used source of information about consumers in traditional markets, especially for frequently purchased products (or FPP such as grocery and clothing), is purchase history (such as recency, frequency and monetary value of customers' transactions). Similar information is scarce for online retailers that are typically faced with visits from relatively anonymous or "unidentifiable" visitors who form a significantly higher proportion of traffic than "loyal" or "registered" customers. This difficulty is especially pronounced for online retailers of less frequently purchased durable goods, who face an interesting scenario. On the one hand, these

retailers usually know very little about the customers that visit their (online) store due to the lack of identifiable historical interactions. At the same time, given the nature of the purchase involving big-ticket and high-involvement goods, customers are more likely to conduct extensive pre-purchase research and place greater value on appropriately targeted information that improves the utility of their purchase (Mack 2009; PriceGrabber Consumer Behavior Report 2009). Faced with limited interactions and a slim dossier on each customer, durable good retailers must seek alternate ways to learn about their customers' needs and preferences. We explore one source of rich micro-level and real-time information contained in the store-level clickstream data available to online retailers.

The goal of our research is twofold. First, we seek to meaningfully characterize consumers' search and navigation behaviors within a session obtained from clickstream data in ways that reflect shopping-relevant underlying differences. We refer to these differences as *states of shopping*. Second, we examine whether product and price-related online information had different impacts on conversion for customers belonging to various *states of shopping*, and whether information varied in its impacts on purchase related behaviors *within a session* (complete purchase) and *across sessions* (return visit and future purchase).

Consumers search online to learn about the assortment of available products, brands and prices across firms. By the time they finish their online research, many will have made up their minds on what specific product and /or seller to buy from. Even when consumers don't buy within a session, they take away useful knowledge about available alternatives – information that is likely to significantly influence their preferences and purchase outcomes later (Mandel and Johnson 2002; PriceGrabber Consumer Behavior

6

Report 2009). Furthermore, several studies report that retailers' web sites trump manufacturer sites and search engines as the information sources cited by consumers as most frequently used for conducting product research online (Compete Online Shopper Intelligence Study 2010; iCrossing Report 2010). It is therefore crucial for firms to both understand the shopping-related needs conveyed by consumers' online search, and act on it by providing the right types of information at the appropriate times to the customer. Fortunately, with the growth of clickstream technologies, there has been a phenomenal improvement in our ability to understand customers. Clickstream data offers the ability to analyze not just the purchase occasion alone, but also the sequence of events that lead to desirable outcomes within a website (Montgomery et al. 2004). Consumers' clicks provide retailers with fine-grained insight ranging from their relative level of interest across categories and their consideration sets, to the types of information accessed and their purchase-related outcomes.

While it is possible for the retailer to provide a variety of purchase-related information, and allow consumers to pick and choose, this is often suboptimal because of concerns involving information overload. For a shopper, a cluttered screen is often a challenge to navigate; and increases the probability that consumers may overlook important pieces of information. On the contrary, consumers may be more receptive to and better served by information that is well targeted to their shopping needs. Prior research has found that a large part of consumers' pre-purchase search activity involves seeking both *price* and *product* information, and both have been found to impact consumers' choices and outcomes in different ways (Diehl et al. 2003; Klein and Ford 2003; Lynch and Ariely 2000). The most commonly sought after price information in

online channels – free shipping and sales or discounts – are no doubt important in helping consumers consummate their purchase; but promotions are frequently margin-eroding (e.g., Gelb et al. 2007). Moreover, it is not clear whether they help build a loyal customer base that will continue to return to the store, or whether consumers that respond to price information are price-sensitive shoppers who seek the lowest price across retailers and are less inclined to return when they leave. If the latter was true, focusing on pricing alone to attract and convert customers may not be the best long term strategy. As online shopping matures consumers are also increasingly seeking to research and understand the available product assortments better. Retailers, in turn, are responding by investing costly dollars in providing a rich multimedia experience for their online customers in an attempt to differentiate from other retailers (Tedeschi 2006). This involves the use of some combination of visual merchandising, product configurators, and buying guides, among others. Whether rich product-related information can turn casual visitors into repeat visitors, and increase their propensity to purchase remains untested. If it did, what types of customers are most likely to benefit from the availability of rich product information? We lack understanding of when consumers value price-related information more than product-related information, and vice versa. Relatedly, are there certain shopping states when product (rather than price) information would help move consumers further along the shopping process and closer to conversion? We explore answers to these questions.

We develop cookie-level panel models to describe and assess the impacts of online information on purchasing both within and across sessions for consumers belonging to different latent states of shopping defined on the basis of their observed session-level behaviors. Our model allows consumers to belong to different states across

sessions, and can therefore capture state transitions through time for a subset of consumers with repeat visits. We estimate our models using clickstream data from a leading click-and-mortar retailer in the U.S. market that covers visits from 77,574 customers to four best-selling durable products carried by the retailer in late 2006. We find that a three-state model comprised of *directed shoppers, deliberating researchers* and *browsers*, best describes the latent shopping-relevant differences across customer sessions in our data. We then uncover important differences in the effects of information across the three types of sessions. When examining the impacts on purchase conversion *within a session*, product information had the strongest impact for deliberating researchers, while price information about a category-level discount proved useful for both directed shoppers and browsers. Price information related to site-wide free shipping had a positive impact across a broad set of sessions, highlighting the value placed on free shipping by consumers who shop online. More surprising were the two negative effects of information that led online customers to delay a purchase or abandon a session. We found that discounts or sales that apply to all products in a given category (e.g., 10% off refrigerators) had a negative effect on deliberating researchers, while rich product information that highlights various features of product alternatives in a category hampered the purchase process of directed shoppers. We describe interesting reasons for these unexpected effects of information.

Our next set of findings highlight important tradeoffs in the effects of product and price information on *within-session conversion* versus two *across session outcomes* - future purchases and the likelihood of repeat visits. Whereas price-related information had positive impacts on within-session conversion for a larger set of sessions, both types

of price information negatively influenced purchase for returning visitors. When online customers did not purchase upon receiving price or promotion information, they were in general less likely to purchase in future sessions if they returned. Additionally, we observe that price information had contrasting effects on customers' within-session conversion and inclination to revisit the store. In contrast, product information positively influenced a smaller set of customers to convert within a session, but had a strong impact on across-session purchase behaviors, influencing consumers to both revisit and purchase in later sessions across all three states of shopping.

The rest of the paper is organized as follows. In §2, we survey existing research and discuss our conceptual framework. We describe the data in §3, and develop our empirical models and strategy for uncovering latent states of shopping in §4. We develop a cookie-panel model and describe our main findings in §5, and examine the robustness of our findings in §6. We conclude with a discussion of the implications in §7.

2.2. Background and Conceptual Framework

2.2.1 Review of relevant literature

Given our interest and goals in this study, we draw from two main streams of literature. The first stream focuses on characterizing consumers' clickstream data as a new source of insight into their shopping needs and intentions, and the second stream focuses on understanding how different types of information affect consumers' purchase-related outcomes. In turn, our study aims to combine these insights to develop a micro-level model of user behavior that can serve as a useful starting point for targeting product and price information in online channels.

In the first stream, a large body of existing work spanning computer science, information systems and marketing has been devoted to studying consumers' search and navigation behaviors in online channels, and broadly suggest that search paths and patterns can predict outcomes. Scholars in CS and IS have examined users' paths or traversals on the web in an attempt to understand how users surf the World Wide Web and to classify their navigation strategies. This literature on web usage mining uses descriptive measures to characterize search into meaningful or 'interesting' patterns (Canter et al. 1985; Catledge and Pitkow 1995; Tauscher and Greenberg 1997; Yang and Padmanabhan 2007; among others) and learn their associations with desired outcomes (Cooley et al.1999; Srivastava et al. 2000). Recently, information systems researchers have begun to incorporate user intention (Jin, Zhou and Mobasher 2004) and contextual information (Adomavicius et al. 2005; Palmisano, Tuzhilin, and Gorgoglione 2007) into the study of user search paths. The resulting models have been then used to implement better document or page pre-fetching systems, recommendation systems and adaptive personalization systems in online environments (e.g., Perkowitz and Etzioni 1998). Other studies have used paths within the context of e-commerce to construct micro-conversion metrics based on look-to-click rate, click-to-basket rate, and basket-to-buy rate (e.g., Gomory et al. 1999) and to compare the navigation patterns of customers to those of non-customers (e.g., Spiliopoulou et al. 1999).

In marketing research, scholars have used path data for predicting conversion likelihood. Some studies have examined paths taken by consumers across websites (e.g., Johnson et al. 2004; Park and Fader 2004); while others - more closely relevant to our study – have focused on search within a website (e.g., Bucklin and Sismeiro 2003;

Montgomery et al. 2004; Moe and Fader 2004; Sismeiro and Bucklin 2004). Among the second set of studies, some have examined search within a session (e.g., Moe 2003; Bucklin and Sismeiro 2003) and others have modeled sessions over time (e.g., Moe and Fader 2004). However, majority of the studies in this literature have either lacked access to or have not modeled the effects of the different types of content (and information) seen by the consumers - which is likely to have significantly influenced a large proportion of behaviors.

Studies in the second stream have examined the impact of price and product information found online on consumers' market outcomes at the aggregate level (e.g., Hodkinson and Keil 2003; Ratchford, Lee, and Talukdar 2003; Viswanathan et al. 2007; Zettelmeyer et al. 2005). However, they typically do not distinguish the impacts of information across different types of consumers. It is possible that certain types of consumers benefit more from product information than price-related information and vice versa. Further, while much of the existing work has examined the final purchase outcome, it is useful for retailers to understand whether and how information impacts other related shopping behaviors such as and adding products to the shopping cart and returning to visit the store.

Our research extends these streams of work to understand how clickstream patterns and behaviors can be used to characterize customer sessions in meaningful ways that not only differ in navigation patterns, but that also distinguishes consumers according to their needs or state of shopping. We then subsequently examine which customer groups (or sessions) respond better to product vs. price information, thus

combining relevant findings from both streams of work to develop guidelines for targeted information provision at a micro-level.

In this vein, our work is closest in spirit to a limited number of existing works that study consumers' responses to marketing communications and/or prescribe strategies to optimally target messages to individual customers or segments. Rossi et al. (1996) studied the problem of using purchase history to design optimal target marketing in the offline market. In recent times, studies in this stream have focused on the impacts of advertising on the web. Zhang and Krishnamurthi (2004) study the related questions of when-how much-and to whom to promote to in an online market for frequently purchased products FPP on the basis of past purchase history. Chatterjee, Hoffman, and Novak (2003) model the click-proneness of consumers or their response to web-based advertising efforts using clickstream data, but do not examine whether clicks led to purchase outcomes. Manchanda et al. (2006) study the effect of banner advertising on purchasing behavior using a limited clickstream dataset. Since they only observe those consumer visits to the site that resulted in a purchase, they estimate a conditional model of the effect of advertising on consumers who buy at least once. Also, they do not observe the content of the advertisements- and therefore cannot distinguish between the impacts of product vs. price-related information and promotions.

In contrast to these existing works, we use a rich clickstream dataset to distinguish among different types of product and price information made available by retailers, and identify the level of exposure to information in an individual session. We then examine the impacts of information by combining techniques from the clickstream modeling of user paths and econometric modeling that allows us to account for both session-level and

cookie-level heterogeneity in unobservables. Another important modeling distinction in our work is the examination of the final purchase outcome, rather than brand-choice which is commonly studied in much of the existing works in the marketing literature (a few exceptions are Manchanda et al. (2006) and Bucklin and Sismeiro2004)). In our study, we therefore fix the brands and focus on the purchase outcome. Finally, by studying the online purchase incidence of durable goods, our works adds complementary knowledge to the literature that primarily focuses on FPP. Specifically we expect that online information would play a greater role for durable good purchases where past experience and experiential learning may be limited, and therefore consumers are likely to conduct extensive pre-purchase search for product and price-related information.

Next, we describe techniques to identify consumers that differ in their shopping orientation, and three types of product and price-related information provided by retailers in online settings.

2.2.2 Consumers' Latent States of Shopping

It is now well understood that not all shoppers are in the same state or mindset when shopping for products, and these underlying differences are known to be reflected in their offline search behaviors (Cox 1967; Putsis and Srinivasan 1994). Such variances are likely to translate into the online market as well. Clickstream data, in particular, are composed of navigation trails from a diverse set of customers, who have varying purchasing needs and goals (Bucklin and Sismeiro 2003; Moe and Fader 2004). Treating consumers as homogenous may thus be erroneous.

In the offline channel, several studies relay empirical support to the ability to use observed behaviors such as decision making strategies to meaningfully infer consumer

differences (e.g., Olshavsky 1985; Payne et al. 1993). However, given the limited ability to track consumers' search offline these models tend to lie at an aggregate level. In this study, we examine consumers' search at a more nuanced level that allows us to examine their intermediate decision-making and information-seeking behaviors, thereby, providing greater insight into both how consumers search and navigate, and what (information) drives their purchase behaviors. We thus infer consumers' latent shopping needs and orientation from their observed session-level behaviors.

We borrow from existing studies that have attempted to characterize these differences in several ways. Li et al. (1999) categorized customers by their online "shopping orientations", but focused on demographic determinants rather than buying behaviors. Koufaris et al. (2001) focused on the type of search mechanisms used online, but did not provide any typology of search behaviors. In a set of studies that closely informs our work, Hoffman and Novak (1996) and Dholakia and Bagozzi (2001) divided online customers' search processes into goal-directed and experiential. These studies however did not discuss the antecedents or the effects of the mind-sets on purchase outcomes. Building on this stream, Moe (2003) divided online customers into four categories based on customers' search behaviors (directed versus exploratory) and purchasing horizon (immediate versus future). Using content viewed online, consumers were categorized as belonging to buying, deliberate-searching, browsing, and knowledge-building states in the shopping process. The existing literature thus suggests that consumers at an online retail store would be likely to fall within the continuum of latent states extending from *exploratory* browsers to *directed* shoppers.

Exploratory browsers are undirected, less-deliberate and stimulus-driven (Janiszewski 1998). This type of search, as found in prior literature, may not necessarily be motivated by a specific goal, and consumers derive utility not from the outcomes of search, rather, from the process of searching and visiting a site. Experiential behaviors are often part of a consumer's ongoing search process (Hoffman and Novak 1996; Wolfinbarger and Gilly 2001). By contrast, directed searchers are focused in their search and are driven by a goal (Janiszewski 1998). Consumers who conduct directed search obtain utility by clicking and traversing through paths that allow them to gather information related to a product of interest or an impending purchase (Childers et al. 2001; Titus and Everett 1995; Hoffman and Novak 1996).

2.2.3 Online Information

The web offers retailers the valuable ability to influence customer purchase behaviors by providing them information in real-time as they browse or shop their online store. While information in online markets may come in a variety of forms, we focus on information that is directly provided by the firm to actively searching consumers at its online store – the content and availability is therefore under the control of the firm[1]. Past research has generally found that both product- and price-related information play key roles in a firm's information provision efforts by providing consumers with appropriate information to aid in the reduction of uncertainty or costs associated with the purchase of products (Diehl, Kornish and Lynch 2003; Klein and Ford 2003; Lynch and Ariely 2000).

---

[1] Additionally, consumer-generated content such as reviews and ratings may also be classified as information available at an online store, when it is made available by the seller on the store website. However, we do not study this type of information since its content is not usually under the control of the retailer. We also do not include advertising information sent to passive consumers (e.g., via email or banners) with the goal to induce them to visit (rather than purchase once in) the online store.

In this study, we specifically consider three types of information- category-specific product information, category-specific price information and generic price information. This information is retailer-provided and generally not brand-specific. *Product information* provides consumers with greater knowledge related to the capabilities, features, uses and applications of the products in a product category, thereby allowing consumers to better "experience" products (Lucas 2001). Such superior product knowledge may help consumers to lower their uncertainty and increase their utility for products in that category. In online markets, such non-price information may include the use of multimedia and microsites to provide rich media product configurators, buying and comparison guides, and video/audio demonstrations of features.

Price information informs consumers about ways to lower the monetary cost or price associated with the purchase of products. We identify two separate types of price information- *category-specific price* information and *generic price information*. Category- specific price information offers consumers price incentives to purchase products from select product categories (such as "Huge savings on home furnishing-10% off", "Tool sale- buy one, get one free", "End of season special values on all kitchen appliances"). Generic price includes information about a price reduction or discount that may be applied to any purchase at the firm's website and is therefore not specific to any one particular category. Examples in this category of price information include offers on shipping and delivery fees (such as "free shipping on orders over $X", or "free shipping today").

From the retailer's point of view category-level product and specific price information helps to increase the attractiveness of all products in a category whereas

generic price information increases the utility for any product in the web store. Examining the impacts of information at this level is consistent with our interest to study purchase incidence rather than brand choice. While not intending to be comprehensive, our categorization of online information captures a bulk of the types of product and price related information that online retailers use today and is a useful starting point for teasing apart the effects of information on different types of customers that visit an online store.

2.2.4 Discussion

We examine whether consumers respond differently to the three types of product and price information provided by the online retailer. We use a data-driven approach to empirically determine the optimal number of states of shopping. Aside from resulting in differences in observed search and navigation behaviors, membership in various latent states is likely to differentially impact the likelihood of purchase (Moe 2003). In the traditional channel, researchers have described the existence of a purchase funnel that consists of a sequence of increasingly directed or focused stages that consumers progress through when making purchase related decisions (see Lee and Ariely's (2006) shopping goals theory; and Trope and Lieberman's (2003) construal level theory among others). We expect to find that on average customers who are browsing will have a lower baseline purchase propensity than customers who are closer to the directed shopper end of the state spectrum and often further ahead in the planned purchase process. Consumers in the middle of this spectrum have in past studies been found to be researching about an impending purchase in a product category of interest (e.g., Moe 2003). We expect that their baseline purchase propensity would therefore lie in between the other two.

Controlling for their baseline purchase propensities, we are interested in the impacts of online price and product information.

Directed buyers have typically completed their research and information gathering process, and are closer to finalizing their purchase. It has been observed that consumers at this stage shop around retailers and price comparison websites to determine the locations of acceptable low prices (Wolfinbarger and Gilly 2001). We expect that such consumers will therefore benefit most from the availability of price promotions than product information because they offer the best value for their already selected product(s). A sale in a product category that a directed consumer wants to make a purchase in or a (free) shipping offer can be extremely successful in incentivizing her to complete the purchase, and preventing her from delaying the purchase, or worse, abandoning the site in search of better deals elsewhere. At the other extreme are consumers who are browsing the retail store and spend their time visiting several product departments or categories. Often, they do not have a particular product purchase in mind when they start their session. Some subset of browsers may also be seeking knowledge about a category that they are interested in but perhaps not considering making a near-time purchase (Moe (2003) refers to them as knowledge builders). Thus we may observe unplanned or impulse purchases from this group when they obtain information that renders a purchase sufficiently attractive. In an industry study sponsored by the Yankee Group and Ernst and Young (2002), the top two factors that contributed to such a spontaneous impulse purchase indicated by survey respondents were a special sale price (75% respondents) and free shipping (49% respondents). Thus purchases made by browsers more likely to result from obtaining price-related information that give the

appearance of a "good deal" or a "value buy". Product-related information, on the other hand, can engage the browsing customer, and help them discover new categories. Some work exists that suggests that spontaneous purchases can be driven by strong emotional reactions to products (Rook, 1987; Rook and Gardner, 1993), or by causing consumers to become more involved in the product category (Bloch and Bruce, 1984; Laurent and Kapferer, 1985; Schmidt and Spreng 1999) – both of which may be evoked using rich product stimuli. The relative effect of product vs. price information on browsers remains to be tested.

Consumers whose state of shopping lies in between directed buying and browsing actively seek to obtain information to learn about available brands/features and decide amongst alternatives. Thus, product information in the form of buying guides, configurators and rich multimedia tools can be useful in educating the consumer and enhancing their product experience, while also helping move them closer to completing the purchase or becoming directed buyers. Specific and generic price-related information, however, do not help consumers make choices among or compare alternatives since they render all alternatives in a product category more attractive.

Finally, we are interested in studying the relative effects of product and price information on purchase oriented outcomes within a session versus across sessions. Next, we describe our data and empirical strategy.

2.3. Data

We use a unique clickstream dataset obtained from a leading click-and-mortar retailer of durable goods. We are interested in purchase incidence, and therefore fix the products of interest. Accordingly, we chose the top four best-selling products at our

retailer (henceforth referred to as *focal products*), and obtained all relevant clickstream

where that product was viewed (clicked on) during consumers' visits to the e-tailer. This

extensive dataset includes all searches conducted by online consumers who visited the

retailer's website and clicked on one of the focal products during a contiguous 30-day

period in late 2006. Each visit by a consumer to the website is recorded as a session,

consisting of an ordered and time-stamped sequence of clicks to the online store pages.

The clickstream data is a rich source of information about consumers' activities at a

website and provides detailed insight into the type of pages viewed including category

pages, product pages, information pages, promotions, customer service, catalogs etc. It

also contains information on consumers' use of various search tools and decision-aids to

refine and screen product alternatives using price, brand, features and other attributes.

See Tables 2.1a and 2.1b for a view of the partial clickstream of two users, one who

doesn't purchase and another that buys in a session, respectively.

Clickstream data offers some benefits over data from websever logs. In the latter

case, each page request by a user tends to generate several server hits from graphics,

multimedia, and content on the page, thus requiring that the hits be aggregated to

correspond to a meaningful user page request. With clickstream, each page view

corresponds to a single individual page or URL requested by the user, making it much

cleaner and more complete. However, clickstream data is in text form and requires

extensive pre-processing before it can be formally analyzed. The information contained

in each click must be parsed in order to determine the nature of the content that the

customer viewed. The clickstream data is first filtered using a custom-built parser written

in the PERL scripting language, which makes sense of the information content of each

click, and encodes the text into numeric form amenable to quantitative analysis. In order to accurately encode the content seen by the consumers, we also downloaded all relevant pages from the retailer's online site during the period of data collection.

2.3.1 Sample Construction

The total number of unique sessions in the dataset, identified by a unique combination of cookie ID and session ID, equals 86,231. We eliminate sessions which included only one page view[2], and also removed sessions where no product pages and products were viewed. An important limitation of using clickstream data is that we cannot determine with certainty what product was purchased if we do not observe the product that the consumers clicked on to add to the shopping cart. We therefore limit our examination to sessions where the consumer clicked on a focal product to view it[3]. This resulted in 43,041 sessions. Finally, to ensure that we do not capture only the repeat sessions of visitors who might have made their first visit in the days preceding our data collection, we dropped all sessions in the first two days of our sample time period. This choice is supported by findings from a study of over 150 million online transactions across 800 retailers that found that when shoppers left an online store due to concerns about security, brand trust, and the need to price-compare, nearly 80% of those who return did so within 1-2 days (McAfee 2009). Our final sample consists of a total of 40,740 sessions from 36,636 unique users (cookies). The total number of sessions that are repeat visits is 4,102 resulting in 7,104 total sessions (17.44%) from 3,002 repeat visitors.

---

[2] These one-page visits could simply comprise store hits where the consumer accidentally landed on the firm's website, specifically a focal product page, from a search engine and immediately abandons the session.
[3] There are a few cases where we observe products added to the shopping cart without the consumer having clicked on the product to view it. This highlights a limitation of clickstream data in that we only have information on actions captured as a click. If a consumer views the product without actually clicking on it, and subsequently adds it to the cart, we cannot know for certain what product was added. To avoid any confounding, we do not include such cart adds in the analysis.

In this session level sample, the visit to cart ratio is 9.31%, the visit to buy ratio is 2.06%, whereas the conditional cart to buy ratio is 22.12%. At the cookie level 2.30% of consumers make a purchase.

2.3.2 Measures

Outcome: The purchase outcome is measured as a binary variable that indicates whether a consumer completed the purchase ($Purchase$). As a first step, we track whether consumers added a product to the shopping cart (measured using a binary variable $Cart$). Following this, consumers at the retailer's store have to complete a series of three steps in order to complete a purchase – these are tracked as step1, step2, and step3 in the clickstream data (see Table 2.1b). A purchase is considered complete only upon completion of step3. Sessions where consumers added to cart, but did not complete step3 are referred to as abandoned sessions. In addition, we also construct a count measure ($Purchase\_cnt$) that is closely related to Purchase – the number of purchase related steps completed by a user during a session, with a higher count indicating greater likelihood of completing the process.

Information: The retailer uses the following features to enhance consumers' product-related experience and to provide product-related information ($ProdInfo$): product buying guides, multimedia to demonstrate product features, and tools to provide dynamic design ideas to promote a product category. All four product categories offered consumers multiple product information features and tools. Price-related information in turn provides the consumer with information on monetary incentives associated with purchasing a product. Generic price information ($GPriceInfo$) includes an offer of free shipping available store-wide for 16 days during our data collection period. Specific price

intervention ($SPriceInfo$)refers to information on the sales and promotions available for products in specific product categories. During the time of the data collection, there were category-specific price sales for three product categories. These three types of information had differing patterns of availability at the store, thereby allowing us to separately identify the effects of each. Further, the retailer did not target the price and non-price promotions to customers. A binary variable indicates whether consumers obtained each type of information - $ProdInfo, GPriceInfo, SPriceInfo.$

States of Shopping: Individual sessions form the basis for categorizing consumers' state of shopping, which as theorized, can change across sessions (over time) for a given consumer. However since the state is actually latent, we infer it from observed search behaviors and navigation patterns of consumers across the website. We borrow from past work in identifying variables including the breadth, depth, and intensity of search to differentiate between directed versus browsing behaviors (e.g., Moe 2003; Wolfinbarger and Gilly 2001).

The breadth of search is defined using two measures - the number of unique product departments ($DeptBreadth$)viewed, and the number of unique product categories viewed[4] ($CatBreadth$). The first refers to search across highly unrelated product categories or departments- e.g., clothes, food and decorations. The second refers to search within a department and across related product categories- e.g., men's clothes, women's clothes, and accessories. The depth of search reflects the extent of hierarchical search within the product category of the focal product ($Depth$), and is measured as the maximum number of times the customer drilled-down or hierarchically narrowed down

---

[4] The online store is broadly organized as departments (Appliances)→ categories (kitchen appliances)→ specific product categories (refrigerators)→ products.

24

the search results. This measure is normalized since the four product categories allow for a different maximum depth by design of the category.

For intensity of search, we create a set of variables that measure the level of involvement the shopper experienced in a given session. We measure the total time spent in minutes[5] ($TotalTime$)and the number of pages ($TotalPages$) visited in the session. We also include squared values of these two measures. We count the total number of unique product pages viewed by the consumer during the session ($TotalProducts$). A product page is counted when the consumer clicks on a particular product to view its details, and not if a product is only seen listed as a collection of available alternatives within a category. Additionally, we create two related ratios –the number of product pages accessed per minute ($ProdPagesPerMin$) with a lower number indicating that the consumer is more engaged with (reading and processing rather than skimming) the content, and the ratio of product pages to the number of categories visited during the session ($RatioProdtoCatPages$)where a larger number would indicate either that the customer was focused and searched only a few categories and/or that she viewed many product pages (across all categories that she accessed).

Controls: We control for several additional variables that may impact purchase outcomes. The first set consists of consumers' use of electronic decision aids to screen and refine the available assortment of products in a category. Past research has observed that by changing the composition of considered alternatives, the use of these tools may shape consumers' decision processes and have significant impacts on consumers'

---

[5] A well-known limitation of clickstream data is that the length of time spent on the last page is not recorded. This information is censored because while we know when the user accessed the last page, we do not know when the user left the site (Bucklin and Sismeiro 2003). However, since duration is only a control here, this is not problematic.

purchase behaviors (see Alba et al. 1997; Diehl et al. 2003; Hoffman and Novak 1996; Lynch and Ariely 2000; Winer et al 1997). We distinguish between consumers whose refining and screening criteria were primarily focused around price related attributes such as "under X dollars", "between X and Y dollars" ($PriceFacetedSearch$) versus product related attributes such as brands and features ($ProdFacetedSearch$)[6]. Another available tool is text-based search, using which consumers can directly search and locate items of interest using a textbox ($TextSearch$). This is an alternative to searching for products by using hierarchical search or drill-down through departments, categories and sub-categories. Finally, consumers in our study were also able to conduct comparisons of selected products using a side-by-side comparison matrix ($CompMatrix$). All four measures are calculated using consumers' extent of tool usage during a given session.

We include date controls ($Date$)in order to account for fluctuations or differences in the online environment from one day to the next that are not observed by us. We include the time of the day of the session is measured using dummies for morning, afternoon, or evening/night ($TimeofDay$), whether the session was started on a weekend ($Weekend$), the month of the session visit as a binary for September or October ($MonthofVisit$), whether the session was a repeat visit ($RepeatVisit$), and the order of the session within a cookie ($OrdSession$). We also include dummy variables to classify the session as having conducted a search for one of the four product categories ($ProductType$), and an indicator for whether the consumer logged into a user account at the website ($Account$) prior to adding a focal product to the cart. We also track the number of times during the session that the consumer viewed the following types of

---

[6] We note that this measure of a consumer's price-product sensitivity is distinct from their responses to product and price related information. Consumers are also allowed to sort, but our clickstream does not capture this information.

pages: home and sitemap related pages (*HomePage*), local retail store and catalog pages (*StorePages*), pages external to the retailer's site (linked to from the retailer's site) (*ExternalPages*), and pages that are marked with error messages (*ErrorPages*). We also measure views of user generated content such as reviews and ratings (*UGCReviews*).

2.4. Session Level Model of Purchase Behavior

As a first step, we are interested in distinguishing sessions or visits made to the online retailer by customers. The notion of the purchase funnel suggests that consumers may progress from relatively undirected to more focused states of buying. In order to allow consumers to belong to different latent states of shopping across multiple sessions, we categorize consumer behavior at the level of a session given only observed values. Our data display a high proportion of zeros, as is expected in a purchase dataset. One approach to handle this is to use a hurdle or zero inflated models that separate the probability of obtaining a zero outcome from the probability of nonzero outcomes (Winkelmann 2008). However, we expect that consumers belonging to any state will experience a non-zero probability of purchase – that is both non zero and non-zero values can be realizations from the same underlying stochastic process. This aspect is better captured using finite mixture models[7]. This is because all consumers who have a zero count of purchase related pages do not necessarily belong to the same subpopulation or distribution – rather, browsers who do not buy have different underlying reasons than directed buyers who do not buy, and the finite mixture model can accommodate that

---

[7] The finite mixture model produced better overall fit than alternative ways to handle the prevalence of zeros. Furthermore we focus on classifying sessions rather than consumers, which in later models will allow consumers to belong to different latent states across visits – as observed in the real world.

therefore producing a better fit for our data. Our underlying belief is that product and price related information found in online environments may have different impacts on purchase likelihood and outcomes for consumers in different states of shopping, which in turn may be proxied through shopping and searching behaviors observed in clickstream data. Thus while we use the observed breadth, depth and intensity measures to help identify consumers belonging to different states, our main interest lies in then examining how the three types of information affect outcomes across states.

In §2.4.1, we adopt a clustering via mixture models approach to determine consumers' latent states of shopping – where it is assumed that data are generated by a mixture of underlying probability distributions in which each component represents a different group or cluster. While our primary dependent variable is *Purchase* – due to complications in the identification of binary mixtures[8], we apply the mixture model to the count outcome measure (there is a high correlation between *Purchase* and *Purchase_cnt* with $\rho = 0.904$, $p = 0.000$) where we model all components as derived from the same distributional family, namely, poisson for which generic identifiability exists (Titterington et al. 1985).

In §2.4.2, we describe the results from the session-level model and determine that a 3-segment solution provides the best fit for categorizing sessions. We then use a highest posterior probability assignment rule to assign sessions to a latent state of shopping. This assignment is then used to re-examine our model and its robustness to how we measure purchase using the binary outcome and a full set of controls.

---

[8] Identification in a binary outcome model generally requires that we observe consumers repetitively in T> 2K-1 sessions (K is the number of component distributions). Unfortunately due to data limitations, that would lead us to drop a large mass of our sample sessions.

2.4.1 Model

Let Y be the non-zero integer valued random variable that measures the count of purchase completion pages visited by the user in a session. In the base-case Poisson regression model, the probability mass function of Y is given by

$$P(Y = y) = \frac{\exp(-\lambda)\lambda^y}{y!}$$

Where λ is the mean or E[Y]. In a Poisson mixture model, λ is treated as a stochastic variable with mixing density function f(λ).

$$P(Y = y) = \int \frac{\exp(-\lambda)\lambda^y}{y!} f(\lambda) d\lambda$$

Further, finite mixture models treat the mixing density as discrete and arising from a fixed number of components G with the probability that an observation belongs to g=G equal to π_g and component specific mean or rate λ_g.

$$P(Y = y) = \sum_{g=1}^{G} \frac{\exp(-\lambda_g)\lambda^{yg}}{y!} \pi_g \tag{1}$$

The log of the component-specific rate is modeled as a linear function of covariates though to exhibit differences across latent subgroups of consumers. Given that mixture models can get easily complicated to estimate when the parameters grow, we estimate a simple yet parsimonious model to determine the usefulness of a mixture setup for our data. The covariates include whether consumers obtained each of the three types of information online, and the set of seven breadth, depth and intensity BDI measurers (RatioProdtoCatPages is excluded due to collinearity).

$$\lambda_{ig} = \lambda_g(x_i, \beta_g) = exp(\beta_g' \, x_i)$$

$$= exp\ (ProdInfo_i'\boldsymbol{\beta}_{prod} + SPriceInfo_i'\boldsymbol{\beta}_{sprice} + GPriceInfo_i'\boldsymbol{\beta}_{gprice} +$$

$$BDI_i'\boldsymbol{\beta}_{bdi})$$

Where $i = 1, ..., N$ sessions   (2)

$\boldsymbol{\beta}_{prod}, \boldsymbol{\beta}_{sprice}, \boldsymbol{\beta}_{gprice}$ are the coefficients of online information

$\boldsymbol{\beta}_{bdi}$ are the coefficients of breadth, depth and intensity measures

As discussed above, we anticipate that consumer sessions belonging to different latent states of shopping may experience varied impacts of covariates, specifically product and price information obtained during the session on purchase outcomes. Such response heterogeneity is well documented in the marketing literature (e.g., Chintagunta 1993; Jain and Vilcassim 1991; Wedel and Kamakura 2000). In the past researchers have observed that there may be unobserved heterogeneity not only in the intercept parameter but also in the slope parameters, i.e. the covariate coefficients (Allenby and Rossi 1999). As shown in prior work by McLachlan and Peel (2000) and Wedel and Kamakura (2000), finite mixture models are a useful tool to segment or group observations by differences in the effects of covariates on the dependent variable. Each support of such a heterogeneity distribution can be interpreted to represent a subset of (consumer) sessions in the online store, and can be used to differentiate among sessions. The finite mixture model uses a discrete mixing distribution of the parameters and simultaneously estimates both consumers' membership in latent states and their *session-level* response parameters to improve both the identification of states and model fit across the states.

It would be appropriate here to briefly discuss alternative ways to model heterogeneity. Finite mixture models have been shown to outperform traditional post hoc approaches involving cluster analysis (Vriens et al., 1996). Random coefficient models

apply a continuous mixing distribution to efficiently estimate average effects but they remain uninformative about responses at a specific disaggregated levels (or segments) which are of interest to us. Hierarchical Bayesian models estimate individual-level parameters but have been shown to be equivalent in performance to finite mixture models in identifying latent heterogeneity across several analyses[9]. Given our interest in identifying the latent states of shopping that groups of customer sessions resemble, we determine that a finite mixture model is especially useful and managerially appealing in our context, and therefore considered more appropriate than other alternatives.

As a middle ground between pooled and individual heterogeneity models, finite mixtures assume that the observed variables come from a population consisting of a finite number of homogeneous groups. We assume that the observations y_i are drawn from a G-component density **f**, and the mixture distribution is given by the weighted sum across the **g** components.

$$\textbf{Pr } (\boldsymbol{y_i} \in \textbf{population g}) = \ \boldsymbol{\pi_g} \tag{3}$$

The g-component mixture density is given by:

$$\boldsymbol{f(y_i| ProdInfo_i, SPriceInfo_i, GPriceInfo_i, BDI_i; \ \Omega_1, ... \Omega_G; \ \pi_1, ..., \pi_G) =}$$

$$\boldsymbol{\sum_{g=1}^{G} \pi_g f_g(\ y_i| ProdInfo_i, SPriceInfo_i, GPriceInfo_i, BDI_i \ ; \ \Omega_g)} \tag{4}$$

$\boldsymbol{\pi_g}$ is the prior probability that observation y_i belongs to component **g. $\boldsymbol{0 < \pi_g < 1}$** and $\sum_{g=1}^{G} \boldsymbol{\pi_g = 1}$. For identification, we follow the labeling restriction that $\boldsymbol{\pi_1 \ \geq \ \pi_2 \geq}$ $\cdots \geq \ \boldsymbol{\pi_G,}$ which can be satisfied by rearrangement after estimation (Titterington et al. 1985)

---

[9] We refer interested readers to discussions in Wedel et al. (1999); Andrews, Ansari and Currim (2002) and others that compare the benefits of using finite mixture models vs. Hierarchical Bayesian methods to model heterogeneity.

Ω_g are component parameters that are estimated by maximizing the following log likelihood

$$\max_{\pi,\Omega} LL =$$

$$\sum_{i=1}^{N}\left(\ln\left(\sum_{g=1}^{G}\pi_g f_g(y_i|\ ProdInfo_i, SPriceInfo_i, GPriceInfo_i, BDI_i;\ \Omega_g)\right)\right) \quad (5)$$

The posterior probability that observation y_i belongs to component **g** given by Bayes theorem conditional on observed covariates and the outcome.

$$\Pr(y_i \in \text{population g} | ProdInfo_i, SPriceInfo_i, GPriceInfo_i, BDI_i, y_i; \Omega) =$$

$$\frac{\pi_g f_g(y_i|\ ProdInfo_i, SPriceInfo_i, GPriceInfo_i, BDI_i, \Omega_g)}{\sum_{g=1}^{G}\pi_g f_g(y_i|\ ProdInfo_i, SPriceInfo_i, GPriceInfo_i, BDI_i, \Omega_g)} \quad (6)$$

Each session is then assigned membership into a group representing a different latent state of shopping for which it has the largest (posterior) probability.

2.4.2 Estimation and Results

The model is estimated using the EM algorithm within the maximum likelihood framework (Dempster, Laird and Rubin 1977). For each fixed value of the number of components, the unobserved component memberships of the observations are treated as missing values and the data are augmented by estimates of the component memberships, i.e. the estimated a-posteriori probabilities, iteratively. Estimation of the model requires the provision of initial values for cluster membership, and we reran the models with several random starting points in order to avoid settling on local optima. Identification of count based mixture models has been proved by Teicher (1963) who show that a necessary condition is that matrix of covariates be of full rank.

We estimated our model by increasing the components from 1 to 4. Prior work suggests that since regularity conditions for the use of Likelihood ratio tests do not hold,

it is more appropriate to use information criteria to select the best model. We use three

commonly used criteria: AIC and AIC3 (Bozdogan 1987), and BIC (Schwarz 1978). The

model is chosen on the principle of parsimony that all else equal, for the same log

likelihood, we should prefer a model with fewer parameters. The best model is the one

that minimizes $-2 * LL + p * d$; where p is the number of free parameters in the model

and $d = 2$ for AIC, $d = log\ N$ for BIC, and $d = 3$ for AIC3. There is recent evidence

that the AIC3 measure is more appropriate for discrete data (Andrews and Currim, 2003),

and has shown remarkable performance in identifying the true model with only minor

overfitting in Monte Carlo studies (e.g., Dias and Vermunt 2007). These studies also

found that increasing the sample size usually led to reduction in the overfitting (except

for AIC). AIC, on the other hand tends to choose the model with more parameter

complexity, while BIC places a heavy penalty on complexity, and for small or moderate

samples, often chooses models that are too simple. In our model, these model based

criterion suggest that the 3-component solution provides the best (or second-best) fit for

our session-level data as indicated in Table 2.2. Moreover, as discussed next, this

clustering provides meaningful groupings of customer sessions that are likely to be useful

for online retailers who lack other information about their customers.

2.4.3 The Latent States: Characterization and Results

Our goal in using mixture models is to uncover underlying differences across

consumers. From a managerial perspective as well, the usefulness of segmentation lies in

its ability to uncover meaningful groupings that obtain different benefits from product

and price related messages. We begin by characterizing differences across the three

obtained states as displayed in Table 2.3a. Sessions in State 1 had the lowest number of

unique department and category breadth – that is they visited very few categories compared to sessions in State 3, which had the highest numbers on both breadth measures. Customers in State 2 performed the highest number of hierarchical drill-downs (depth of search starting from departments to categories to products) while customer sessions in State 3 contained the fewest.

Next, we assess an array of variables that indicate the level or intensity of focus displayed by a customer in their session. Customer sessions categorized as State 1 viewed the highest number of pages and spent the longest time on the website. Sessions in State 2 and 3 differed little along these two attributes. However, customers in State 2 viewed a significantly higher number of product pages (nearly double that of customers in state 1). Thus while customers in State 1 viewed more pages overall, only a small share were product pages, and a majority included pages related to the store, promotions, specials, and retailer policies. Another related distinguishing variable is the ratio of product level to category level pages which is the highest for sessions in State 2, followed by sessions in State1 and then State 3. This variable provides one measure of the intensity of product search conducted within (focal) product categories. For sessions in State 3 this lower number indicates either that they viewed fewer product pages or conducted a dispersed search across many categories. Customers in State 2 and State 1 viewed significantly fewer product pages per minute than browsers did indicating that the former may have spent more focused time engaging with (reading about) products. Finally, State 1 sessions had the highest likelihood of being a repeat visit for a cookie.

On the basis of the above characterization, we conclude that sessions in State 1 resemble directed buyers who are the most focused in their search and purchase activities

(Moe 2003), sessions in State 2 are similar to searching and deliberating users who are conducting research and learning about products in the focal category, and sessions in State 3 are best described as browsers or experiential window shoppers whose interests were not focused. We refer to these three states of shopping as ***directed shoppers*** (DS), ***deliberating researchers*** (DR) and ***browsers*** (BR) henceforth. While not necessarily perfect, the categorization highlights the most prominent behaviors observed across these clusters[10].

Next, we examine the purchase outcomes associated with membership across states, which we expect to differ if our labeling of the states was reasonable. We find that sessions from directed shoppers had the highest overall proportion of users that both added products to the cart (14.49%) and completed the purchase (5.13%), whereas sessions from browsers had the lowest overall proportions for both purchase related behaviors. Interestingly, we observe that conditional on adding a focal product to the shopping cart, browsers had a higher likelihood of completing the purchase (20.2%) than deliberating researchers (17.2%), but lower than directed shoppers (35.4%). Further, upon examination of customers' use of tools to refine and screen alternatives, we find that customers conducting research were the least likely group to use such decision aids, indicating their greater reliance on compensatory choice processes in building their consideration sets. Whereas, directed shoppers and browsers displayed greater non-compensatory search through the use of decision aids to quickly narrow down the available assortment. Directed shoppers displayed a high usage of text search to directly

---

[10] Our dataset is limited to customers who visited and/or purchased one of four focal products. In characterizing the sessions, we examine the nature of their search behaviors across departments and categories to measure the extent to which their search was focused or dispersed. Thus, irrespective of whether these customers purchased a focal product or another product, their classification would still correctly describe their search behaviors.

find products they wanted, and also were most likely to use the product comparison matrix which allows users to side-by-side compare up to 4 chosen products.

In related past work, Gupta and Chintagunta (1994) used demographic variables to determine consumers' segment membership in an offline context. However, with the availability of micro-level search behaviors, we believe that using measures such as breadth, depth and intensity of online search during a visit /session will better describe consumers' latent heterogeneity in shopping needs and goals. Moreover, using the variables in BDI_i provides (online retailers) an actionable strategy to profile segments since these variables are measurable using clickstream data whereas demographic variables are typically unavailable to online durable good retailers.

Next, we examine the coefficients of information on the number of purchase completion pages visited by the user in a session. These coefficients are of primary interest in our study and are presented in Table 2.3b. In column (1) are the results obtained from the mixture model that uses the count of purchase completion pages visited by the user as the outcome. We also estimate additional session-level models using the categorization of sessions obtained from mixture modeling and a full set of controls including month, time of day, product type, counts of various types of pages viewed (error, store, home etc) and use of tools and decision aids (facets, comparison matrices, UGC) as described earlier. In column (2) the outcome is $Purchase\_cnt$ while in column (3) it is $Purchase$. The main results that we observe in Table 2.3b are the following. Product information had the strongest positive impact on purchase outcomes for deliberating researchers, followed by browsers, and had little to a negative impact on directed shoppers. Generic price information displayed a strong influence on all three

types of customers – with the largest impact on directed shoppers followed by browsers. Specific price information had positive impacts on directed shoppers and browsers, while it had a negative influence on deliberating researchers.

Finally, we assess the transitions between the latent states uncovered by the mixture model for cookies with multiple or return visits as displayed in Table 2.4. We observe a high level of inertia for directed buyers (66%) and information gatherers (63%), where consumers are likely to continue in the same state. For browsers, the likelihood of returning as a browser is close to 50% and as a deliberating researcher is 41%.

2.5. Cookie level model of purchase behavior

In this section, we develop a random-effects cookie-level panel model that allows us to examine both the *within-session* and *across-session* influence of online product and price information. Within-session refers to the impact of information obtained in a session on purchasing in the same session, while across-session refers to the impact of information obtained in a session on purchasing (and purchase-relevant behaviors) in future session(s). The panel specification allows us to account for two forms of heterogeneity. The first is cookie-level unobserved heterogeneity which is stable within a cookie and time-invariant across its sessions, modeled using random effects. Additionally sessions from a cookie may belong to different latent states of shopping across repeat visits. This time-variant heterogeneity is modeled using session-level dummies to represent the state following the categorization determined in §2.4.

While it is possible to allow all covariates to differ in their impacts on the propensity to purchase across the three latent states of shopping, we are only primarily interested in the effects of online information. Therefore, we focus on characterizing these varied impacts of online information through the use of interactions between them and the three states- DS, DR, and BR. We also include dummies to represent the state of the preceding session (if any). This allows us to track consumers as they change their state over time and examine the effects of state transition patterns on purchase outcomes. Lastly, the across-session impacts of online information are estimated from consumers who make repeat visits to the retailers' website. We include measures of accumulated exposure to price and product information (across past sessions that are observed by us) and examine their impacts purchase outcomes in the current visit.

Next, we describe the model setup for our primary outcome – the binary measure *Purchase*.

2.5.1 Model

$$y_{it}^* = PastProdInfo_{it}' \, \gamma_{pProd} + PastSPriceInfo_{it}' \, \gamma_{pSprice}$$

$$+ PastGPriceInfo_{it}' \gamma_{pGprice} + PastCart_{it}' \gamma_{pCart}$$

$$+ ProdInfo_{it}' \, \gamma_{prod} + SPriceInfo_{it}' \, \gamma_{sprice} + GPriceInfo_{it}' \, \gamma_{gprice}$$

$$+ LATSTATE_{it-1}' \, \gamma_{pLatstate} + LATSTATE_{it}' \, \gamma_{latstate}$$

$$+ \left[ProdInfo_{it}' + SPriceInfo_{it}' + GPriceInfo_{it}' \right]$$

$$* LATSTATE_{it}' \, \gamma_{info*latstate} \quad + BDI_{it}' \, \gamma_{bdi} + \, x_{it}' \, \gamma_x + \, \alpha_i$$

$$+ \, \eta_{it}$$

$$where \; i = 1, \ldots \ldots, N \; \; cookies \; \; t = 1, \ldots. T \; sessions \qquad\qquad (7)$$

$$y_{it} = \mathbf{1}(y_{it}^* > 0)$$

$$v_{it} = \alpha_i + \eta_{it}$$

$\gamma_{pProd}, \gamma_{pSprice}, \gamma_{pGprice}$ are the coefficients of online information accumulated in past

sessions

γ_{pCart} is the coefficient of the number of cart adds in past sessions

$\gamma_{prod}, \gamma_{sprice}, \gamma_{gprice}$ are the coefficients of online information obtained in the current

session

$\gamma_{pLatstate}$ are the dummies that represent the (latent) state of the immediately previous

session

$\gamma_{latstate}$ are the dummies to represent the (latent) state of the current session

$\gamma_{info*latstate}$ are the coefficients for the interactions between information and latent state

γ_{bdi} are the coefficients of the breadth, depth and intensity variables

γ_x are the coefficients for observed session-level control variables including month, time

of day, product type, counts of various types of pages viewed (error, store, home etc) and

use of tools and decision aids (facets, comparison matrices, UGC) as described earlier.

α_i is the unobserved cookie-level individual random effect which is assumed to be

uncorrelated with the covariates. These individual effects are distributed $\alpha_i \sim N(\mathbf{0}, \sigma_\alpha^2)$.

η_{it} is the i.i.d. random error term; $\eta_{it} \sim N\,(\mathbf{0}, \sigma_\eta^2)$ and represents unobservables that are

uncorrelated across sessions and cookies. The variance of v_{it} is given by $\boldsymbol{Var}(v_{it}) =$

$\sigma_v^2 = \sigma_\alpha^2 + \sigma_\eta^2$ and $\boldsymbol{cov}(v_{it},\ v_{is}) = \sigma_\alpha^2$ if the sessions belong to the same cookie or

consumer (irrespective of the time lag between sessions), and 0 otherwise. The variance

of the pure shocks is normalized to one. The fraction of the total error variance due to the

individual consumer level component of the error term is given by the intragroup

correlation coefficient

$$\rho = \frac{\sigma_\alpha^2}{\sigma_v^2} = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\eta^2} = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + 1} \tag{8}$$

Let z contain the covariates in equation (8), then the probability of observing the given

outcome conditional on the cookie random effect is given by

$$\mathbf{Pr}(y_{it} \mid z_{it}, \alpha_i) = \Phi\left[(z_{it}'\gamma + \alpha_i)(2y_{it} - 1)\right] \tag{9}$$

The likelihood for each unit given is given below

$$L_i = \mathbf{Pr}(y_{i1}, y_{i2,\dots,}y_{iT}) = \int_{-\infty}^{\infty} \phi\,(v_{i1}, v_{i2}, \dots v_{iT} \mid \alpha_i)\,\phi(\,\alpha_i)\,\partial\alpha_i \tag{10}$$

Since the dependence between v_{it}'s is attributable to the shared variation in α_i (due to the

assumed independence between z_{it} and α_i), the need to integrate across a T-variate

normal distribution is eliminated. By conditioning on α_i, we integrate them out of the

likelihood and evaluate the one-dimensional integral in (10) by using Gauss-Hermite

quadrature (Greene 1997, p.190). The Log-likelihood of the model described in (8)-(10)

is given by:

$$LL = \sum_i \sum_t y_{it}\,ln\,F(z_{it}'\gamma) + (1 - y_{it})\,\ln\,(1 - F(z_{it}'\gamma)) \tag{11}$$

2.5.2 Results

The results from the cookie-panel model are presented in Table 2.5. We include

dummy variables to capture the effects of sessions belonging to one of the three latent

states of shopping. In columns (1)-(3), we present the coefficient estimates for the partial

models with step-wise additions of the interaction terms between the three states of

shopping and product/price information to assess if there are interactions among

information types. The full model is presented in column (4), and is the one we use. We discuss the results from Tables 2.5-2.6 below[11].

States of shopping: The mean purchase likelihood for directed shoppers in the sample is given by the constant in Table 2.5, and by a linear combination of the constant and the respective latent state dummy for deliberating researchers ($LatState\_DR$) and browsers ($LatState\_BR$) when all other covariates are held at their mean (or median for binary variables). That purchase is a rare event in our data is reflected in the negative coefficients for all three states of shopping, with the highest rate for directed buyers and the lowest rate for browsers. On average, we note that directed shoppers had the highest rate of conversion at 5.12%, followed by 1.80% for researchers and 1.29% for browsers, controlling for covariates.

State transitions: Next, we examine the impact of state transitions across sessions for repeat visitors. In this analysis, the baseline consists of sessions without a past state - they are the first visit for that cookie in our data. As seen in Table 2.5, the coefficient of $PastLatState\_DS$ is negative, while the coefficients of $PastLatState\_DR$ and $PastLatState\_BR$ are positive and significant. Thus, returning to shop after being in a directed buying state had a significant negative effect, while returning to shop after having been in either deliberating or browsing states had significant positive effects on the likelihood of purchase. Sessions abandoned by directed shoppers are thus a costly loss - when these customers leave without purchasing, their likelihood of doing so when they return is significantly lowered. This result suggests that retailers should focus on trying to convert directed shoppers in the current session itself. In Table 2.6, we examine further

---

[11] We also ran a fixed effect model which was limited only to cookies with multiple sessions. We obtain results that are qualitatively consistent with those obtained here. These are available upon request.

details to help understand the impacts of transitioning between latent states of shopping. We find that among directed shoppers that do return, the likelihood to complete the purchase drops sharply. If they return as researchers, this conversion rate is 4.57%, whereas it drops to 3.52% and 2.70% respectively when they return as directed shoppers and browsers. On the other hand, when consumers' transition into the directed state of shopping after being in the other two states the results are optimistic. For instance, for sessions where consumers transition from deliberation and research to directed shopping, the purchase likelihood jumps to 34.83% and for sessions where consumers proceed to directed buying after browsing, this number improves to 29.41%. Overall, for deliberating researchers and browsers we find that transitions to the directed state of buying had the highest likelihood of converting in the next session, followed by transitioning to researching and last, browsing.

Past online information: We turn our attention to the impacts of online information accumulated from past (but not the current) sessions or visits to the retailer. In Table 2.5, the coefficient of $PastProdInfo$ is positive and significant, while the coefficients of $PastSPriceInfo$ and $PastGPriceInfo$ are negative, with only the effect of $PastSPriceInfo$ significant. Among the three types of information, the cumulative effects of product information obtained in earlier sessions had a positive impact on a consumers' likelihood of purchasing in a given session. By contrast, the accumulated effects of category specific price promotions and to a lesser extent generic price promotions obtained in the past sessions had a negative effect on purchase in a given session. This finding highlights the potential negative future effects of promotions when consumers expect them but they may no longer be available.

Current online information: To examine the contemporaneous or within-session purchase impacts of information obtained by a consumer during the session itself, we need to assess not only the three main-effect coefficients of information but also the interactions with consumers belonging to different latent states of shopping. For ease of understanding and comparison, these coefficients are calculated from the estimates displayed in column (4) in Table 2.5 and provided in Table 2.7.

Product information ($ProdInfo$) had the strongest impact on within-session purchase for deliberating researchers, followed by browsers. Consumers who are conducting research, gathering information and deliberating about a product category are the ones that display the greatest positive response to the information contained in product buying and use guides, how-to documents, and multimedia demonstrations of product features. This provides empirical confirmation of an intuitive result. Product information provides the information necessary to assess and compare the products available within a category, thereby allowing customers conducting researching for an impending purchase to form their preferences. The impact on browsers is interesting. We find that product information had a positive impact on customers who were not necessarily focused on the particular category, suggesting that such information may have attracted customers to a product category. Browsing customers who may have had a general interest in the product category but not necessarily considering a near-term purchase and received product information appeared to purchase more often than browsers who did not receive product information.

In contrast to the impact on deliberating and browsing customers, product information appeared to lower the likelihood of purchase in a given session when

43

presented to directed buyers. The negative effect on directed buyers is surprising. One possible explanation is that receiving detailed product information at this stage creates ambivalence or distraction when such information contradicts consumers' original impressions or preferences, especially if it highlights product-relevant aspects that the consumer may have overlooked or ignored before. Some past works have found that under certain circumstances, the use of decision tools and recommendation agents in online settings may provide suggestions that are counter to the preferences of users, thereby causing negative reactance (e.g., Fitzsimons and Lehmann 2004). We expect that such reactions might have caused directed buyers to delay (or abandon) their purchase upon obtaining product related information. Unlike browsers who have most likely not engaged in behaviors that create a commitment to purchase, directed buyers have in the recent past invested active time and effort (perhaps elsewhere) in considering the purchase. Thus obtaining product related information in the form of help and buying guides, and multimedia demonstrations appears to distract the latter type of buyer, while it attracts the former.

Category specific price information ($SPriceInfo$) had significant positive impacts on both directed shoppers and browsers, leading them to convert more often than in its absence. Consumers who display directed behaviors at a website are typically highly focused on a product category, and have usually completed their product research and have narrowed down their consideration sets and are not seeking more product-related information (Moe 2003). Such consumers may price-shop across retailers as they look for deals or promotions on the specific product(s) that they are considering. Obtaining relevant promotion-related information can therefore incentivize them to

purchase from the said retailer. Browsers, on the other hand, are likely to convert upon receiving promotion related information if they consider the purchase price to be an attractive deal. Unlike directed buyers, browsers displayed a broader interest across product departments and categories during their session – suggesting that while they were interested in the focal product and its category, they may not have been actively seeking out related information. Obtaining price information on an attractive promotion or sale in a focal product category may therefore serve to generate interest and influence consumers to respond with an impulse purchase. In some cases, an impulse purchase may be driven because the browsing customer encounters information that stimulates their memory and reminds them about a product(s) that he/she had planned long before to purchase but had postponed or delayed it while awaiting to gather more information (perhaps about sales).

Interestingly, specific price information did not induce similar effects on deliberating researchers, and had a negative effect on their purchase behavior. This appears counter-intuitive at first, but to see why recall that these consumers are still conducting research and deliberating about and forming consideration sets. Specific price information increases the attractiveness of all products within a focal product category, and does not change or alter the relative attractiveness of product alternatives. Obtaining information about a category price promotion improves the valuation of all products in a category, but this increased attractiveness of products might also mean that more alternatives now satisfy the feasibility constraints of a shopper. Thus, rather than help the customer move closer to making a purchase; in fact specific price information may delay their decision-making by increasing the number of alternatives whose (sale price adjusted) values are now acceptable.

Consider a shopper with a budget constraint, who is still researching product alternatives (and has not determined a desired product). Learning about an x% off discount in a focal category might now additionally make other alternatives attractive that were deemed too expensive in the absence of such a promotion, thereby increasing the choice or consideration set. Such a process might lead consumers to inaction or deferral of decisions, and has been observed in a variety of laboratory settings (e.g.,Chernev 2003; Dhar 1997; Dhar and Simonson 2003; Gourville and Soman 2005; Iyengar and Lepper 2000) and real world online settings (e.g., Nunes and Boatwright 2001). Broadly these studies find that consumers delay a purchase when required to negotiate difficult trade-offs between alternatives. Adding more alternatives to the choice set caused choice overload, increased choice conflict and resulted in choice deferral. The deferral was observed to be greater when the assortment considered by the consumer was increased to include alternatives that were non-alignable (Chernev 2003; Gourville and Soman 2005). This is likely to happen when, for example, more appliances fall into a consumer's feasible set of alternatives, but they include machines that vary in the availability of features or attributes, thereby making comparison among them more difficult for the consumer. Thus, encountering larger selections can actually reduce purchases within a given product category (Gourville and Soman 2005). Overall, this finding suggests that for consumers in the deliberation and research state, price or promotion related information by itself is insufficient to motivate them to complete the purchase. They are instead likely to continue researching and gathering valuable knowledge about the products in the focal category.

Generic price information ($GPriceInfo$) or promotions related to shipping fees tended to have across-the-board positive impacts on consumer sessions belonging to all three states, suggesting that shipping offers continue to be highly valued by online buyers. In other words, the absence of free shipping or related promotions appears to lower the purchase likelihood for all consumers. This result is also supported by recent studies conducted by PayPal and comScore that found that the leading cause of shopping cart abandonment cited by 46% of respondent was high shipping charges[12]. The strongest positive impact of shipping related price offers is interestingly observed for browsers, followed by directed shoppers and then researchers. Directed shoppers are those who are close to finalizing their purchase and have a deeper commitment to the purchase than browsers. This result suggests that directed buyers are less likely than browsers to abandon their purchase when a shipping offer is not available. Researchers who have not yet completed their evaluations and formed their preferences are only weakly (nevertheless significantly) influenced by free shipping offers.

2.5.3 Additional Analyses Using a Restricted Sample

In order to assess the validity of our results, we reran our model using an additional restricted sample as follows. One limitation of using clickstream to study consumer's purchase outcomes is that we cannot ascertain the true intent or motivation of consumers. While consumers may have visited a product page sometime during the session, it may not translate into true interest in the product and need not suggest that the product was considered for purchase by the consumer. We therefore place a stronger restriction on the customers whose sessions will be included. In this second sample, we

---

[12] Eighth Annual Merchant Survey (April 2009) sponsored by PayPal and comScore.

require the customer to have displayed "substantial interest" in one of the focal products during at least one of his/her visits to the store. We consider the act of adding a focal product to the shopping cart as an indication that the customer is interested in the product, and therefore include all sessions from this customer. For other customers who did not add a focal product to the shopping cart at any time during our data collection period, we include sessions from only those customers who viewed the focal products multiple times during at least one of his/her visits to the store (in most cases the focal product was the last product to be viewed before the session was abandoned). This results in a sample of 11,408 sessions from 8,842 unique cookies where some consumers viewed and added the focal product to cart, whereas other consumers viewed the focal product and did not add to cart but rather abandoned the session afterward. We refer to this as the *interested sample* (as opposed to the *full sample*), and display the results from the panel purchase model in column (5) in Table 2.5. The coefficients of the effects of the information across the states of shopping for this restricted sample are calculated and displayed in column (2) in Table 7. These results are broadly consistent with column (1) in Table 2.7, and engender confidence that our main findings about the effects of information on the purchase outcome.

2.5.4 Tradeoffs Between the Within-Session and Across-Session Impacts of Information

Together, the results of the influence of three types of information underscore an important observation – that there is a tradeoff between the effects of product and price related information on purchase outcomes within a session and across sessions. Information about products in a focal category aid consumers in researching product

alternatives, learning about product features, uses and applications, and has significant within-session influence on purchase behaviors for deliberating researchers (and to a smaller extent browsers in some models). But it negatively influences the within-session conversion of directed shoppers. Exposure to product related information, however, had the strongest positive influence on purchase decisions for returning consumers irrespective of their state of shopping in previous sessions. On the contrary, both types of price information displayed strong positive effects on within-session purchase behaviors for customers in two latent states of shopping, but had weak to strong negative impacts on purchase for returning customers who abandoned sessions previously. We explore this tradeoff further below.

In addition to the above demonstrated contrasting effects on purchase that product and price information obtained in the past have, we examine whether online information influenced *consumers who do not purchase to return to visit the online store* in the future. In Table 2.8, we model the likelihood of a session visitor's likelihood of returning to visit as a function of the online information received in the current session. We specify a panel model to control for cookie-level unobservables, and the state of shopping is modeled using dummies and interactions as before.

For directed shoppers - the group of customers with the highest conversion rate- we find the coefficient of price information about discounts in a specific product category ($SPriceInfo$) to be negative. This suggests an interesting tradeoff for $SPriceInfo$, whereby it has a significant positive effect on helping directed shoppers to complete a purchase within a session, but when such a customer does not purchase and leaves (perhaps, in search of better deals or prices), she is also less likely to return. A similar

pattern of tradeoff effects is observed for shipping related information $GPriceInfo$ on directed shoppers.

Deliberating researchers and browsers however experienced a different pattern of effects of $SPriceInfo$ and $GPriceInfo$ on the propensity to return. In Table 2.8, we see that the coefficients of $SPriceInfo$ are significantly positive for both, while the coefficients for $GPriceInfo$ are insignificant. Thus, we observe a tradeoff between the within-session and across-session impacts of $SPriceInfo$ for deliberating researchers, but in a direction opposite to that experienced by directed shoppers. Promotion information about a specific product category was not useful in converting deliberating researchers into purchasers within a given session (in fact it had a negative effect), but it increased their likelihood of returning to visit the store. For browsers, $SPriceInfo$ had a positive effect on both buying within a session and returning to visit the store. Finally, while $GPriceInfo$ had across-the-board positive effects on within-session purchase behaviors of consumers, it failed to have an effect on influencing abandoning researchers and browsers to return to the store.

In contrast to these effects of $SPriceInfo$ and $GPriceInfo$, our results suggest that customers belonging to all three states of shopping who obtain and view product related information are more likely to return to visit the retailer after they abandon the session. This finding is relevant because it highlights the value of $ProdInfo$ in helping engage the customer and in building a relationship with them that extends beyond a given session. Given concerns echoed by several retailers about consumers who are price-sensitive and respond only to price promotions but are typically not loyal and hunt for deals (e.g., McWilliams 2004), our results show that retailers can benefit by investing in

creating a rich product experience for their customers. Product related information helps

to attract consumers back to the online store, and also has significant impacts on helping

deliberating researchers (and to a lesser extent browsers in some models) to convert

within the session. Thus, our study has uncovered some interesting patterns of effects of

online information on purchase-related behaviors within a session (purchase now) and

across sessions (likelihood to return and purchase in future).

2.6. Robustness Checks

In this section, we conduct additional tests to assess the robustness of our main

findings related to purchase outcomes (as displayed in Table 2.7) to alternate

specifications and explanations.

2.6.1 Endogeneity in Product Information

$SPriceInfo$ and $GPriceInfo$ are available to all customers in our dataset who

view the focal product category on the days that the related discount and shipping offers

were provided by the retailer. However, the impact of $ProdInfo$ on purchase outcomes

may suffer from endogeneity bias if consumers who are more likely to purchase were

also the ones more likely to seek and obtain product-related information. As a first step,

we compare the means or the proportion of customers in each of the three states that

obtained information (see Table 2.3). We observe that fewer deliberating researchers,

who appear to have the strongest positive impact on purchase from $ProdInfo$, obtained

product information than directed shoppers, suggesting that endogeneity may not be a

concern. Yet, in order to more rigorously address the potential for reverse causality, or

the possibility that consumers may self-select or choose to obtain product information for

51

reasons that are also correlated with their purchase outcome, we use the matching method to estimate the effects of $ProdInfo$.

The literature on treatment effects defines the treatment effect of a binary treatment as the difference in outcome when units (here sessions) are treated (receive $ProdInfo$) and when those same units are not treated. However, we only observe sessions in either the treated or the non-treated condition, and therefore must construct the necessary missing counterfactuals for the sessions. Propensity score matching allows us to estimate average treatment effects by comparing the outcomes of treated and control groups that have been matched on the breadth, depth and intensity covariates instrumental in determining the likelihood of receiving treatment[13]. We construct a stratified or matched sample of observations that consists of treated and control groups that are balanced across these observed covariates – and therefore, on average observationally identical. The propensity score is the conditional probability of receiving the treatment rather than being part of the control group given the relevant observed covariates W (Rosenbaum & Rubin, 1983). It is estimated using a probit model as follows where the treatment is $ProdInfo = 1$ and W contains variables that describe breadth, depth and intensity of search.

$$P(\boldsymbol{ProdInfo = 1|W}) = \ \Phi\{\boldsymbol{h(W)}\} \text{ where } \Phi \text{ is the normal c.d.f.} \tag{12}$$

Matching on such a score serves to simulate random assignment of treatment when two conditions hold: a) the observed covariates used to construct the score are balanced, and b) there is no bias from unobserved covariates. We check that condition a) holds, and we restrict the matching to be performed over the common support region – that is using observations whose propensity scores belongs to the intersection of the

---

[13] Additionally, W is chosen to satisfy the Balancing Hypothesis of matching estimators.

supports of the propensity scores of the treated and control sessions. Condition b) is the

Conditional Independence or Unconfoundedness Assumption that treatment assignment

is ignorable (independent of the potential binary outcomes for purchase Y(0) or not Y(1)

in a session) conditional on observed covariates - a critical assumption in matching

models (Abadie and Imbens 2002).

$$P(ProdInfo = 1|W, Y(0), Y(1)) = P(ProdInfo|W) \qquad (13)$$

Identification is achieved when the probability of assignment of treatment is bounded

away from zero and one, known as the Overlap assumption (Abadie and Imbens 2002):

$$0 < P(ProdInfo = 1|W) < 1$$

When these regularity conditions hold, then imbalances in pretreatment covariate levels

can be controlled by adjusting the unidimensional propensity score calculated in (12)

such that comparisons of outcomes occur between treated and control groups that differ

only in their exposure to treatment (Rosenbaum and Rubin 1983). The treatment effect

for an individual TE_i is given by

$$TE_i = Purchase_i (ProdInfo_i = 1) - Purchase_i (ProdInfo_i = 0) \qquad (14)$$

Then aggregate impact of product information on outcomes is calculated as the

sample average treatment effect on the treated ($SATT$) given by

$$SATT = \frac{1}{n_T} \sum_{i \in T} TE_i \qquad (15)$$

Where $n_T = \sum_{i=1}^{n} T_i$ is the number of *treated* units for whom the observed

treatment $ProdInfo_i = 1$. An important concern in using propensity score matching

methods to estimate treatment effects is the potential violation of condition b) above.

While the model accounts for selection on observables, consumers' choice to visit online

product information pages such as buying guides is likely to covary with important

unobservables in the study. The results in §3 and §4 suggest that the varied effects of several relevant variables (breadth, depth and intensity of search and navigation behaviors) on the purchase likelihood are summarily captured in the latent state of shopping. If the latent state simultaneously affects assignment into treatment and the outcome variable, a *hidden bias* might arise to which matching estimators are not robust (Rosenbaum 2002). Additionally, given our interest in separately identifying the effect of (product) information on outcomes across consumer sessions belonging to different latent states, we construct propensity score matching estimates for each latent group separately, in effect using a latent state dummy as a matching covariate in addition to W. This provides us with one way, albeit imperfect, in which to account for unobservables.

The results from the propensity score matching analyses limited to consumer sessions with a common support are presented in Table 2.9. In column (1), matches are found using a caliper or radius matching $(r = 0.1)$, while in column (2), matching is conducted using a block-stratified matching algorithm. The standard errors are calculated using bootstrapping procedures. As observed there, our primary results remain robust. The coefficient of $ProdInfo$ is positive and significant for researchers and browsers, whereas for directed shoppers, it continues to be negative to insignificant.

2.6.2 Price vs. Brand Sensitivity of Consumers

In §5, the results indicated that consumers belonging to different (latent) states of shopping obtained varying benefits from the three types of product and price information. In this subsection, we examine an alternate explanation for the consumer purchase behaviors observed there that we attribute to states of shopping. Were consumers who

completed the purchase when provided with price related (vs. product related) promotions merely more price-sensitive (or brand/feature-sensitive)?

An important feature of shopping online is the availability of refining and screening tools which offer consumers the ability to alter the products that they see, and thereby affect the consideration sets that they build, and the final products that they choose to buy. Prior research has found that such decision tools and aids available in computer mediated markets can have significant effects on the final choices made by customers (e.g., Alba et al. 1997; Haubl and Trifts 2000; Lynch and Ariely 2000). Several retailers (including ours) today provide faceted search tools that lets users refine or navigate a collection of products by using a number of discrete attributes or facets. We are specifically interested in consumers' use of price ($PriceFacetedSearch$) vs. product ($ProdFacetedSearch$) attributes to screen alternatives. We use this as a proxy for consumers' price vis-à-vis product sensitivity for purchases in the focal product category, and examine whether it influenced the results obtained in Table 2.6. If this were the case, we should expect to see that deliberating researchers are more product-sensitive than directed buyers; and that directed buyers and browsers are more price-sensitive than deliberating researchers.

We compare the extent of refining and screening (counts) performed by consumers during sessions (see Table 2.3). We find that on average, deliberating researchers had the lowest counts of product/brand refining ($\mu= 0.172$, s.d. $= 1.148$), followed by browsers ($\mu = 0.371$, s.d. $=1.756$) and directed buyers ($\mu = 0.570$, s.d. $=2.879$). Deliberating researchers also had the fewest number of price refining counts on average ($\mu = 0.250$, s.d. $=1.413$), while browsers ($\mu = 0.555$, s.d. $=2.135$) and directed

buyers ($\mu = 0.579$, s.d. $= 2.843$) had a similar average. Also see figure 2a and 2b for the distribution plots.

Browsers conducted more price-based than feature/brand-based refining and screening operations. Directed shoppers were equally likely to refine using both types of attributes. As a group, researchers were least likely to use either refining criterion (but also relied slightly more on price-based screening). In the results displayed in Table 2.5, we controlled for the extent of price vs. product refining conducted by a consumer in a session. However, neither type of refining significantly influenced consumers' likelihood to purchase, whereas the coefficients for the states of shopping and information were significant[14]. In order to assess whether the three latent states were masking consumer's price sensitivity, we include interaction terms between the three types of information and both types of refining and screening to additionally separate and control for their effects. The relevant coefficients for the three types of online information are displayed in Table 2.10. After controlling for several controls, and the interactions between states and the two types of refining/screening, we find that our main results for the effects of information obtained within a session and in the past sessions on conversion within the session remain consistent with our findings from Table 2.7.

These observations help mitigate the concern that the influence of product and price information merely coincide with corresponding product-price sensitivity of consumers in the focal product category. While price-sensitivity appeared to explain some of the findings related to the effect of specific and generic price information on shoppers, after controlling for the former, the latent state of shopping that the consumer

---

[14] However, we cannot entirely rule out this possibility since the extent of product and price –based refining and screening are only used proxies for, and it may be that these measures do not capture the true underlying sensitivities of consumers.

belonged to continued to remain significant and determined whether product or price related information influenced shoppers to complete their purchase.

2.6.3 Adding to the Shopping Cart

In this subsection, we extend our analysis to examine the impacts of information on an important intermediate or pre-purchase outcome- adding products to the shopping cart. This analysis seeks to shed light on whether product and price information influence shopping cart abandonment- which is a common woe of online retailers (c.f., Murthi and Sarkar 2003) - differently for the three states.

We found evidence of such a behavior in Table 2.4 earlier when we described the states and observed different conditional rates of purchase. We examine this more carefully here in Table 2.11 using a panel model for both the full sample (col 1) and the interested sample (col 2). The impacts of information on adding products to the cart are jointly estimated but separately displayed for each group – the results are largely consistent with our purchase model in §5. Deliberating researchers were more likely to add a product to the shopping cart upon retrieving relevant product information; whereas both directed shoppers and browsers were more likely to do so when they received either type of price-related (sales and shipping) information. This result is interesting because it suggests that the same type of information influences customers in a given shopping state to both add the product to the shopping cart and complete the purchase. This is counter to the belief that once customers have added products to the shopping cart, only price information about promotions and free shipping will influence them to consummate the purchase.

57

We assess the relationship between the two outcomes *Purchase* and *Cart* using a bivarite model that allows us to jointly estimate the effects of covariates across these two. The mean correlation between the standard errors across the two outcomes is estimated to be 0.9935 ($\chi^2(1) = 304.98, p = 0.00$) - this high number indicates that there is high level of similarity in the unobservables that affect a consumer's decision to perform both outcomes. We calculate the predicted probabilities after controlling for the distribution and impacts of several relevant covariates (as used in table 2.5). At a joint predicted probability $\Pr(Purchase_i = 1, Cart_i = 1)$ of 4.44%, the conversion rate is the highest for directed buyers followed by information gatherers (1.73%) and browsers (1.30%). The groups were ranked in the same order for the marginal predicted probabilities of both outcomes - adding to the shopping cart and completing the purchase. However the conditional probability $\Pr(Purchase_i = 1 \mid Cart_i = 1) = \frac{\Pr(Purchase_i=1, Cart_i=1)}{\Pr(Cart_i=1)}$ tells a different story. Conditional on having added products to the shopping cart, directed buyers had the highest probability of completing the purchase (35.35%), while information gatherers had the lowest (18.80%). This suggests that consumers across the different segments perhaps use the shopping cart for different reasons. Deliberating researchers, who add to the cart at a comparatively higher rate than browsers, are however less likely to complete the purchase. The low conditional rate of conversion of shopping carts for researchers underscores the importance of recognizing that some consumers may not be ready to purchase in the current session even if they add products to their cart. They may be using the cart to conveniently hold and compare chosen alternatives as they conduct research and gather more information about products.

2.7. Conclusion

2.7.1 Discussion

We began this study with the goal of determining how firms and retailers should manage the provision of online price and product-related information to customers who are actively visiting their online store. We specifically examined whether customers to an online retail store were distinguishable by their observable search and navigation behaviors accessed through the clickstream that they generate. Following the derivation of a segmentation of customer sessions, more appropriately termed states of shopping in our study, we assessed whether three types of commonly available information differently influenced purchase outcomes across the states.

Our main results are the following. When focusing on *conversion within a session*, both browsers and directed shoppers are best influenced by price-related information (discounts, sales, free shipping etc.). However, customers who are deliberating and conducting research responded best to product-related information. In our sample, we observed that the sessions where customers were deliberating formed the largest group, slightly greater than sessions where the customers were browsing and nearly three times larger than the sessions where the customers were directly buying. This suggests that online retailers have a large potential ability to induce online customers to convert using rich product information if they are able to identify and target the customer when he or she is deliberating and researching the available alternatives in a focal product category. By persuading deliberating researchers to complete the purchase within a session, the retailer reduces the need to have to attract them using price levers when they return later as directed shoppers (or browsers). This allows the retailer to then offer sales

59

and free shipping offers to the customers in states that obtain the greatest value from price-related information, and more importantly might have abandoned the session in their absence. Thus, by uncovering the unobserved state of the shopper, the retailer can appropriately target price vs. product information to the customer, thereby avoiding the need to always offer margin-eroding price promotions in order to incentivize customers to complete the purchase. In fact, our results highlight the surprising negative effect of category-level price promotions on deliberating researchers. In the other negative effect of information, we observed that rich product information distracted directed shoppers and anticlimactically led them to delay their purchase. Thus, our within-session results shed light on the varied impacts of information across customers and also draw attention to the possible undesired consequences of mis-targeted information.

When examining conversion and *purchase-related behaviors across sessions,* our study suggests that there may be important tradeoffs in the impacts of information on purchasing within a session as compared to influencing customers to return to purchase from the online store in a future session. Irrespective of the shopping state of the customer, product related information had a significant positive impact on influencing customers who did not purchase in a given session to both return to the store (in the short-term) and buy (that particular item) in a future session. Our results highlight the important role for product information and its ability to create stickiness in the website and loyalty among its customers.

However, both types of price related information – that had a positive impact on within-session conversion - appeared to have unfavorable or negative impacts on the likelihood of future purchase for directed shoppers. More specifically, when directed

shoppers receive price or promotion information, but fail to find price sufficiently attractive to purchase and therefore abandon the session, they are less likely to return to the store in the short-term for that particular item. Price information on discounts in a specific product category, however, appeared to have positive effects on the likelihood to return to visit (but not necessarily buy) for deliberating researchers and browsers. Thus specific price information might have aided customers to progress farther along the shopping cycle. Finally, free shipping – that had broad positive impacts on within-session conversion for all three states of shopping - failed to have a positive impact on influencing customers in all three states to revisit the store.

These tradeoffs in the effects of product vs. price information within and across sessions is an important finding that our model uncovers due to our ability to not only link customers across sessions but more importantly, track the accumulation of content that they view as they make multiple visits to the online retailer. This allows us to tease apart the effects of information obtained within a given session from the effects of information obtained in past visits. Our findings have relevant implications for online firms, which we discuss next.

2.7.2 Implications

While firms have traditionally had limited and often static opportunities to interact with consumers, the fast-changing environment of electronic retailing is essentially changing this. The availability of micro-level consumer behavior data promises to bring online retailers closer to achieving truly customized interactions with their customers (Alba et al. 1997; Ansari and Mela 2003; Hoffman and Novak 1996). Our study and its findings provide firms with the knowledge that can be a useful starting

point for segmenting sessions from relatively anonymous customers in meaningful ways, and determining the optimal provision of product and price information to these different types of customers. In the absence of identifying information that is typically available in offline channels and for frequently purchase goods, durable good retailers have to devise alternate ways to distinguish their customers. A particularly interesting aspect of our study is the use of observed and easily available search and navigation activity on the website itself to generate the background covariates required to determine the latent shopping state of the customer.

Our study questions the current common practice of offering promotions such as free shipping and product category discounts to all customers that are visiting a store, and provide empirical evidence to support this intuition. We argue that this strategy is suboptimal and results in retailers providing unnecessary promotions to customers who would have purchased anyway. We show that by learning about customers' latent states of shopping, retailers can instead optimally target product and price information to customers who are less likely to complete a purchase in the absence of such information, thereby increasing the lift created by online information.

Moreover, depending on the retailer's goal – immediate conversion in the short term, i.e. before the customer ends a session, versus ensuring that the customer develops a longer-term relationship with the retailer and returns to the site over time – a different information provision strategy is likely to be optimal. This implication is driven by the tradeoffs or contrasting effects generated by our model for product and price information on purchase related behaviors within and across sessions.

2.7.3 Contributions

Our study makes a few important contributions to practice. First, by using in-session or real-time segmentation and customization strategies it allows retailers to avoid the pitfalls surrounding the use of sensitive information about consumers that need to be tracked over long periods of time. In the offline channel, retailers have relied on the use of demographics (e.g., moms vs. teenagers) and purchase histories (e.g., loyals vs. first-timers) to segment customers. This limited retailers to develop information targeting strategies that were based on static customer characteristics and/or past outcomes, whereas, more relevant targeting can be achieved by the use of real-time customer behaviors. This allows us to partially overcome the problem of the "gift-shopper" who is offered irrelevant promotions for children's toys when she later tries to search for business apparel, for instance. Real-time customization strategies enable retailers to better match consumers' concurrent preferences and lead to positive sales outcomes.

Second, since historical actions and pre-determined profiles are not always needed, these techniques may allow retailers to actively target and interact with even *new* visitors to their web store. Third, our model of targeted information is consistent with shifting emphasis from the "static" user model to the "dynamic" behavior model which allows for the same consumer to be targeted in different ways on different occasions based on changing needs/preferences.

2.7.4 Limitations and Future Extensions

Our study adds to a growing stream of research that suggests ways in which firms can improve their customer's online experience by making websites more usable and navigable (Agarwal and Venkatesh 2002; Palmer 2002; Venkatesh and Agarwal 2006),

63

and retailers can aid in consumers' online search and purchase decisions (Novak, Hoffman, and Yung 2000; Hauser 2009). Along these lines, our study sheds light on the impacts of product and price related information for consumers in different shopping states. Our current work is based on a sample observed over a short period that precludes us from studying purchases that may have occurred from customers returning beyond our observation period. We also group together different kinds of rich product information in this work, but it would be useful to tease apart the different effects of buying guides vs. other multimedia demonstrations, for instance. This study should also be extended to study the effects of user generated content such as reviews that is becoming wildly popular in online shopping contexts.

A modeling limitation of our current study is the separation of the tasks of identifying latent states at the session level and estimation of information effects using a cookie-panel. While combining them would require us to make several additional assumptions about the distribution of unknown parameters (that drive the latent state and state transitions) that may not necessarily be realistic, it can help validate the robustness of our current findings. In future studies, it will be useful to examine the pathways of influence – how product vs. price information differently affects customers' underlying purchase oriented structural parameters. For example, what is the impact of information on the buying threshold? Relatedly, when information does not incentivize customers to buy, does it help them to progress through the shopping funnel (and advance from being a browser to a deliberating researcher to a directed shopper)? Finally, while our current work is focused on the impacts of information obtained any time during the session, knowledge about timing or when in the session to provide different types of information

would be complementary, and help firms make even more specific decisions related to optimal provision of online information.

Table 2.1a. Partial clickstream from a sample user who doesn't purchase

CookieID	SessionID	Date/timestamp	Page
4.00E+22	1.16E+27	10/15/2006 16:52	CATEGORY: CATEGORY>GARDEN CENTER-
4.00E+22	1.16E+27	10/15/2006 16:53	CATEGORY: CATEGORY>GARDEN CENTER>CHIPPERS SHREDDERS ACCESS.-
4.00E+22	1.16E+27	10/15/2006 16:53	PRODUCT: 22 IN. 14.4 VOLT CORDLESS HEDGE HOG HEDGE TRIMMER (100060602)
4.00E+22	1.16E+27	10/15/2006 16:53	CATEGORY: CATEGORY>GARDEN CENTER-
4.00E+22	1.16E+27	10/15/2006 16:54	CATEGORY: CATEGORY>GARDEN CENTER>POWER TOOLS-
4.00E+22	1.16E+27	10/15/2006 17:00	PRODUCT: 200 MPH BLOWER VAC (100055950)
4.00E+22	1.16E+27	10/15/2006 17:03	HOME PAGE
4.00E+22	1.16E+27	10/15/2006 17:03	CATEGORY: SUPERFEATURES2/MISCELLANEOUS/PM_FALL_CLEANUP_06
4.00E+22	1.16E+27	10/15/2006 17:04	CATEGORY: SUPERFEATURES2/OUTDOOR_POWER_EQUIPMENT/KH_BLOWERS_BUYING_GUIDE
4.00E+22	1.16E+27	10/15/2006 17:08	CATEGORY: CATEGORY>GARDEN CENTER-

Table 2.1b. Partial clickstream from a sample user who completes a purchase

CookieID	SessionID	Date/timestamp	Page
4.01E+22	1.16E+27	10/15/2006 15:23	SEARCH:BASIC
4.01E+22	1.16E+27	10/15/2006 15:23	CATEGORY: TEXT SEARCH >PATIO SET-CATEGORY>OUTDOOR LIVING-
4.01E+22	1.16E+27	10/15/2006 15:24	CATEGORY: CATEGORY>OUTDOOR LIVING>PATIO FURNITURE>
4.01E+22	1.16E+27	10/15/2006 15:24	CATEGORY: CATEGORY>OUTDOOR LIVING>PATIO FURNITURE>
4.01E+22	1.16E+27	10/15/2006 15:24	CATEGORY: CATEGORY>OUTDOOR LIVING>PATIO FURNITURE - PRICE>$400 - 600-
4.01E+22	1.16E+27	10/15/2006 15:25	PRODUCT: ST. CROIX 5 PC. TILE TOP CHAT GROUP IN FOSSIL (100399316)
4.01E+22	1.16E+27	10/15/2006 15:25	CATEGORY: CATEGORY>OUTDOOR LIVING>PATIO FURNITURE>
4.01E+22	1.16E+27	10/15/2006 15:25	CATEGORY: CATEGORY>OUTDOOR LIVING>PATIO FURNITURE> BRAND>HAMPTON BAY-
4.01E+22	1.16E+27	10/15/2006 15:25	PRODUCT: MISSION BAY 5-PIECE ALUMINUM DINING SET (100397582)
4.01E+22	1.16E+27	10/15/2006 15:25	ITEM ADDED TO CART
4.01E+22	1.16E+27	10/15/2006 15:28	SHOP_CART/PG_ALT_VIEW_POPUP.JSP
4.01E+22	1.16E+27	10/15/2006 15:30	SHOP_CART/PG_DELIVERY_STEP1.JSP
4.01E+22	1.16E+27	10/15/2006 15:32	SHOP_CART/PG_DELIVERY_STEP2.JSP
4.01E+22	1.16E+27	10/15/2006 15:34	SHOP_CART/PG_DELIVERY_STEP3.JSP

Table 2.2. Examining fit across multi-component models (count outcome)

# components	LL	AIC	AIC3	BIC
1	-8221.305	16466.61	16478.61	16569.99
2	-8135.561	16321.12	16346.12	16536.49
3	-8044.895	16165.79	16203.79	16493.15
4	-8032.992	16167.98	16218.98	16607.34

Table 2.3a Describing the latent states of shopping

	State 1 14.32%		State 2 43.18%		State 3 42.53%	
Variable	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.
BDI: Used to predict states						
DeptBreadth	0.337	0.600	0.415	0.602	0.869	0.807
CatBreadth	1.526	2.949	2.448	4.327	4.369	5.688
Depth	2.198	1.205	3.330	1.142	2.094	0.871
TotalPages	30.222	34.326	15.050	14.756	18.692	18.069
TotalTime	18.435	18.019	9.258	11.884	7.093	8.964
TotalProducts	4.614	6.283	8.933	3.973	3.180	3.714
RatioProdtoCatPages	0.794	0.627	0.915	0.962	0.405	0.201
ProdPagesPerMin	2.118	1.730	2.100	1.621	3.179	1.996
Repeat session	0.153	0.360	0.110	0.312	0.074	0.262
Cart	0.145	0.352	0.105	0.306	0.064	0.244
Buy	0.051	0.221	0.018	0.133	0.013	0.113
Conditional Buy	0.354	0.478	0.172	0.378	0.202	0.402
PriceFacetedSearch	0.579	2.843	0.250	1.413	0.555	2.135
ProdFacetedSearch	0.570	2.879	0.172	1.148	0.371	1.756
TextSearch	1.524	4.683	0.853	2.789	0.333	1.751
CompMatrix	0.375	1.604	0.207	1.374	0.131	0.888
ProdInfo	0.160	0.366	0.098	0.297	0.071	0.257
SPriceInfo	0.310	0.462	0.284	0.451	0.263	0.440
GPriceInfo	0.507	0.500	0.504	0.500	0.503	0.500

Table 2.3b Impacts of information on purchase outcomes in session-level models

| Model | (1) Purchase_cnt | | (2) Purchase_cnt | | (3) Purchase | |
| Controls | Basic | | Extended | | Extended | |
	β	s.e.	*β*	s.e.	*β*	s.e.
Directed shopper						
ProdInfo	-0.007	0.189	-0.156	0.289	-0.396***	0.111
SPriceInfo	0.974***	0.260	1.259***	0.331	0.508***	0.136
GPriceInfo	1.885***	0.256	1.081*	0.431	0.732***	0.191
Deliberating researcher						
ProdInfo	1.609***	0.125	1.766***	0.164	0.447***	0.079
SPriceInfo	-0.014	0.223	-0.071	0.286	-0.351*	0.142
GPriceInfo	1.471***	0.186	0.503	0.381	0.465**	0.168
Browsers						
ProdInfo	0.925***	0.209	1.236***	0.248	0.233*	0.095
SPriceInfo	1.258***	0.340	1.247**	0.383	0.988***	0.208
GPriceInfo	2.667***	0.299	1.485***	0.472	1.264***	0.240

Note: The dependent variable is Purchase_cnt in columns (1) and (2) and binary Purchase in column (3).
We estimate session level models with cluster robust standard errors.
*** $p<0.001$, **$p<0.01$, *$p<0.05$, + $p<0.1$

Table 2.4. State transitions for repeat visitors (excluding last session)

	Directed shopper	Deliberating researcher	Browser	Total
Directed shopper	**65.93%**	21.79%	12.28%	904
Deliberating researcher	9.60%	**63.34%**	27.06%	1855
Browser	8.86%	41.25%	**49.88%**	1343
Total	21.77%	46.95%	31.28%	4102

Table 2.5. Estimating the within-session and across-session impacts of online information on completing a purchase
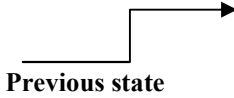
	(1) Full sample β	s.e.	(2) Full Sample β	s.e.	(3) Full sample β	s.e.	(4) Full sample β	s.e.	(5) Interested sample β	s.e.
LatState_DR	-0.566***	0.055	-0.280***	0.057	-0.540***	0.081	-0.126	0.147	-0.119	0.178
LatState_BR	-0.612***	0.062	-0.491***	0.066	-0.501+	0.087	-1.025***	0.225	-1.328***	0.263
PastLatState_DS	-0.349***	0.107	-0.333**	0.108	-0.340**	0.109	-0.323**	0.109	-0.445***	0.122
PastLatState_DR	0.196*	0.084	0.207*	0.084	0.199*	0.084	0.212*	0.085	0.019+	0.093
PastLatState_BR	0.203*	0.089	0.205*	0.090	0.195*	0.089	0.225*	0.090	0.111+	0.099
PastProdInfo	0.177***	0.040	0.184***	0.039	0.180***	0.039	0.180***	0.040	0.175***	0.043
PastSPriceInfo	-0.392***	0.095	-0.408***	0.095	-0.402***	0.095	-0.401***	0.095	-0.371***	0.099
PastGPriceInfo	-0.085+	0.048	-0.115*	0.048	-0.094*	0.047	-0.115*	0.048	-0.128*	0.051
PastCart	-0.217***	0.059	-0.235***	0.059	-0.220***	0.059	-0.235***	0.060	-0.250***	0.066
ProdInfo	-0.344***	0.102	0.204***	0.057	0.201***	0.057	-0.352***	0.103	-0.465***	0.122
SPriceInfo	0.271**	0.099	0.440***	0.114	0.250*	0.098	0.753***	0.148	0.997***	0.179
GPriceInfo	0.672***	0.156	0.716***	0.158	0.604***	0.169	0.802***	0.197	0.932***	0.243
ProdInfo*DR	0.792***	0.119					0.821***	0.120	0.868***	0.143
ProdInfo*BR	0.664***	0.134					0.672***	0.135	0.739***	0.161
SPriceInfo*DR			-0.650***	0.118			-1.012***	0.179	-1.282***	0.211
SPriceInfo*BR			0.053	0.108			0.433+	0.238	0.498+	0.277
GPriceInfo*DR					0.202*	0.098	-0.365*	0.155	-0.534**	0.188
GPriceInfo*BR					0.033	0.102	0.440+	0.229	0.539*	0.267
CatBreadth	0.003	0.018	0.002	0.018	0.001	0.018	0.005	0.018	0.053*	0.022
DeptBreadth	-0.161***	0.025	-0.169***	0.025	-0.166***	0.025	-0.166***	0.025	-0.244***	0.030
Depth	0.150***	0.026	0.158***	0.026	0.157***	0.026	0.150***	0.026	0.257***	0.030
TotalPages	0.104*	0.042	0.105*	0.042	0.101*	0.042	0.110**	0.042	0.043	0.053
TotalPages^2	0.000	0.000	0.000+	0.000	0.000	0.000	0.000	0.000	0.000	0.000
TotalTime	0.287***	0.031	0.296***	0.031	0.300***	0.030	0.279***	0.031	0.235***	0.038
TotalTime^2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
TotalProducts	0.307***	0.049	0.324***	0.049	0.313***	0.048	0.324***	0.049		
ProdPagesPerMin	-0.054	0.040	-0.055	0.040	-0.056	0.040	-0.052	0.040	-0.114**	0.051
OrdSession	0.517***	0.071	0.517***	0.070	0.509***	0.070	0.527***	0.071	0.482***	0.077
OrdSession^2	-0.048***	0.009	-0.046***	0.009	-0.047***	0.009	-0.047***	0.009	-0.040***	0.009
PriceFacetedSearch	-0.006	0.016	-0.008	0.016	-0.006	0.016	-0.008	0.016	-0.004	0.020
ProdFacetedSearch	0.019	0.017	0.019	0.017	0.021	0.017	0.016	0.017	0.017	0.020
TextSearch	0.016	0.013	0.013	0.013	0.013	0.013	0.016	0.014	0.016	0.016
CompMatrix	0.069***	0.021	0.066**	0.021	0.066**	0.021	0.069***	0.021	0.122***	0.026

	(1)		(2)		(3)		(4)		(5)	
HomePage	0.035	0.039	0.033	0.039	0.035	0.039	0.031	0.039	0.094+	0.050
StorePages	0.071***	0.020	0.064***	0.020	0.063***	0.020	0.073***	0.020	0.093***	0.024
ExternalPages	0.104	0.065	0.105	0.064	0.100	0.065	0.113+	0.065	0.160+	0.083
UGCReviews	0.304**	0.107	0.311**	0.107	0.315**	0.106	0.299**	0.107	0.373**	0.138
AccountPages	0.174***	0.032	0.176***	0.032	0.174***	0.032	0.177***	0.032	0.166***	0.038
ErrorPages	-0.195***	0.053	-0.185***	0.053	-0.186***	0.053	-0.197***	0.053	-0.215***	0.062
Intercept	-2.542***	0.388	-2.748***	0.389	-2.605***	0.388	-2.725***	0.404	-0.998*	0.507
σ_α	0.518	0.031	0.517	0.031	0.518	0.031	0.517	0.031	0.537	0.046
ρ	0.212	0.020	0.211	0.020	0.212	0.020	0.211	0.020	0.224	0.030

Note: The dependent variable across all models is Purchase. We estimate cookie-panel models with cluster robust standard errors. Col (1)-(4) use the full sample, while col (5) uses the interested sample.
*** p<0.001, **p<0.01, *p<0.05, + p<0.1

Table 2.6. The conditional effects of state and state transitions on completing a purchase

Previous state → Current state	Directed shopper	Deliberating researcher	Browser
Directed shopper	3.52%	4.57%	2.70%
Deliberating researcher	34.83%	2.72%	4.18%
Browser	29.41%	4.33%	3.88%
Overall	5.12%	1.80%	1.29%

Table 2.7. Impacts of information obtained within a session on completing a purchase

Information obtained	(1) Full sample		(2) Interested sample	
	β	s.e.	β	s.e.
Directed shopper				
ProdInfo	-0.352***	0.103	-0.465***	0.122
SPriceInfo	0.753***	0.148	0.997***	0.179
GPriceInfo	0.802***	0.197	0.932***	0.243
Deliberating researcher				
ProdInfo	0.469***	0.077	0.402***	0.091
SPriceInfo	-0.259*	0.127	-0.285+	0.149
GPriceInfo	0.437*	0.169	0.398+	0.214
Browsers				
ProdInfo	0.319***	0.098	0.274*	0.115
SPriceInfo	1.187***	0.203	1.495***	0.236
GPriceInfo	1.242***	0.238	1.471***	0.286

Note: The dependent variable is Purchase. We estimate cookie-panel models with cluster robust standard errors. Col(1)uses the full sample, while col (2) uses the interested sample.
*** $p<0.001$, **$p<0.01$, *$p<0.05$, + $p<0.1$

**Table 2.8. Impact of information obtained within a session on return visit for non-purchasers**

	(2) Full sample		(2) Interested sample	
	β	s.e.	β	s.e.
Directed shopper				
ProdInfo	0.434***	0.082	0.368**	0.143
SPriceInfo	-1.532***	0.072	-1.464***	0.130
GPriceInfo	-1.836***	0.105	-2.362***	0.247
Deliberating researcher				
ProdInfo	0.295***	0.078	0.297*	0.136
SPriceInfo	0.186**	0.060	0.346***	0.098
GPriceInfo	-0.166	0.094	-0.616**	0.230
Browsers				
ProdInfo	0.168*	0.080	0.160*	0.071
SPriceInfo	0.123*	0.057	0.225*	0.099
GPriceInfo	-0.066	0.094	-0.509*	0.229
σ_α	0.243		0.352	
ρ	0.056		0.110	

Note: The dependent variable is likelihood to return visit. We estimate cookie-panel models with cluster robust standard errors. Col (1) uses the full sample, while col (2) uses the interested sample. The models contain the full set of covariates shown in Table 2.5. *** $p<0.001$, **$p<0.01$, *$p<0.05$

Table 2.9. The impact of product information using matching techniques

	N treated	N control	Average treatment effect	s.e.	T
Radius matching					
All	3883	2389	0.026	0.004	7.094
Directed shopper	932	502	-0.014	0.009	-1.687
Deliberating researcher	1724	1085	0.041	0.006	6.950
Browser	1227	782	0.021	0.006	3.816
Stratified matching					
All	3883	36853	0.032	0.004	8.559
Directed shopper	932	4506	-0.025	0.014	-1.796
Deliberating researcher	1724	14525	0.048	0.006	8.234
Browser	1227	15921	0.026	0.005	4.828
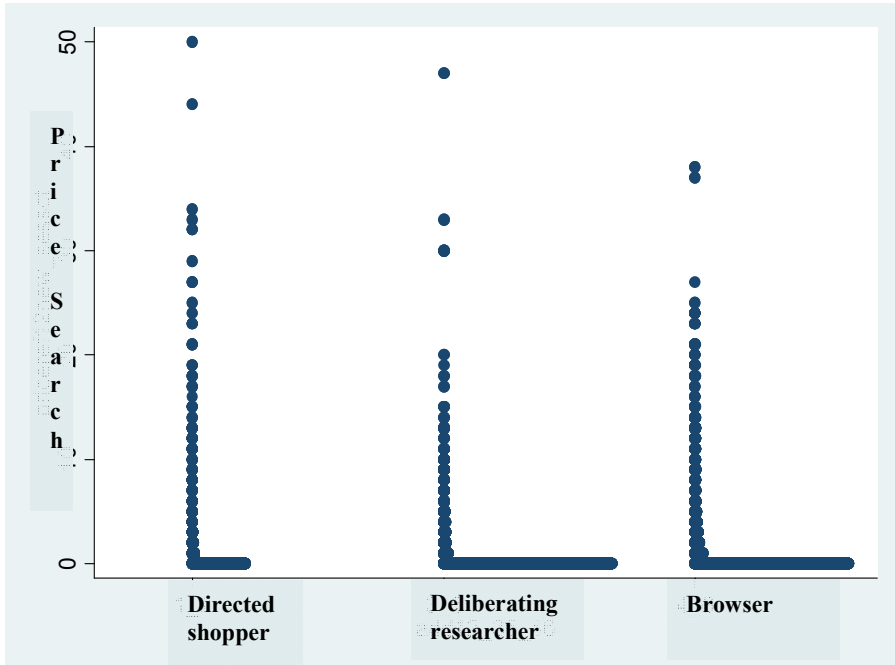
Figure 2.1a. Distribution of the use of PriceFacetedSearch (count) across latent states
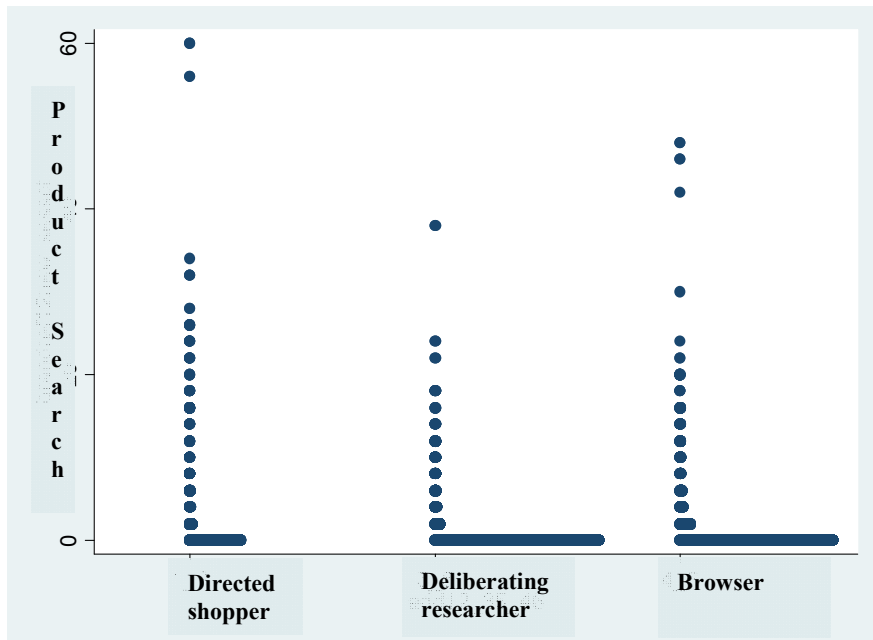
Figure 2.1b. Distribution of the use of ProdFacetedSearch (count) across latent states

Table 2.10. Impacts of information on purchase controlling for price-product sensitivity

Information obtained	(1) Full sample	
	β	s.e.
PAST SESSIONS		
Past_ProdInfo	0.179***	0.040
Past_SPriceInfo	-0.391***	0.095
Past_GPriceInfo	-0.122*	0.048
Past_AddtoCart	-0.235***	0.060
WITHIN SESSION: Directed shopper		
ProdInfo	-0.306*	0.127
SPriceInfo	0.559**	0.178
GPriceInfo	0.301**	0.108
WITHIN SESSION: Deliberating researcher		
ProdInfo	0.684**	0.051
SPriceInfo	-0.456**	0.168
GPriceInfo	-0.092	0.193
WITHIN SESSION: Browser		
ProdInfo	0.080	0.069
SPriceInfo	1.043***	0.222
GPriceInfo	0.812**	0.260

Note: The dependent variable is Purchase. We estimate cookie-panel models with cluster robust standard errors. The models contain the three types of information, past latent state, current latent state and their interactions with information, the two types of price and product faceted search and their interactions with latent states. *** $p<0.001$, **$p<0.01$, *$p<0.05$

Table 2.11. Impacts of information on adding to the shopping cart

Information obtained	(1) Full sample		(2) Interested sample	
	β	s.e.	β	s.e.
PAST SESSIONS				
Past_ProdInfo	-0.024	0.032	0.028	0.035
Past_SPriceInfo	-0.190**	0.062	-0.151*	0.067
Past_GPriceInfo	-0.212***	0.039	-0.152***	0.042
Past_AddtoCart	0.428***	0.046	0.220***	0.050
WITHIN SESSION: Directed shopper				
ProdInfo	0.041	0.072	-0.258**	0.098
SPriceInfo	0.493***	0.149	0.664**	0.211
GPriceInfo	0.285**	0.097	0.427**	0.139
WITHIN SESSION: Deliberating researcher				
ProdInfo	0.684***	0.051	0.639***	0.069
SPriceInfo	0.208	0.131	0.340+	0.190
GPriceInfo	0.038	0.065	0.140	0.097
WITHIN SESSION: Browsers				
ProdInfo	0.080	0.069	0.017	0.091
SPriceInfo	0.649***	0.140	0.939***	0.204
GPriceInfo	0.567***	0.083	0.851***	0.126

Note: The dependent variable is a binary indicating whether the customer added a focal product to the shopping cart during the session. We estimate cookie-panel models with cluster robust standard errors. The models contain the full set of covariates shown in Table 5. *** $p<0.001$, **$p<0.01$, *$p<0.05$

Chapter 3: The Impact of Online Information on the Value of Certification

3.1 Introduction

Over the recent years the growth of the Internet has dramatically increased the information available to prospective buyers across a number of markets. In contrast to traditional settings where information was mostly obtained from a few centralized/institutional sources (typically the seller or a third-party), consumers today have access to information from a multitude of avenues. The Web allows consumers to avail of product and pricing-related information in greater detail and depth, while also providing ways to seek distributed advice from experts and information intermediaries. In addition, consumers have access to user generated content, online word of mouth, product reviews/recommendations, and seller ratings – sources that have garnered appreciable interest recently. Consumers benefit from this rich diversity of online information as they engage in pre-purchase search across several product categories on the Internet (Schadler and Golvin 2005), with over 50% of online consumers reporting that their purchase decisions were significantly influenced by online content (iProspect Report 2007). While it is widely acknowledged that the growing decentralization of information and increased access to such collective intelligence will bring about fundamental transformations in the way firms and consumers transact, there have been few systematic studies examining the implications of these changes for consumers as well as marketers. Our study seeks to address this issue by examining how different types of online information obtained by consumers affects their value for certification in a market with significant information asymmetry.

As is well known, markets with significant information asymmetries - particularly markets for used goods - have traditionally resorted to quality-signaling mechanisms such as certification, warranties, brand, and seller reputation (Dewally and Ederington 2003), to help reduce frictions and the likelihood of market collapse (Akerlof 1970). Such quality signals have been considerably valuable for consumers in these markets, with consumers often paying a premium for them. However, the value of the quality signal to consumers, and the competitive advantage it provides to sellers, depends crucially on the nature and extent of information asymmetries present in the market. With the growth of purchase-related websites, it is possible that the use of product- and price-related online information may alter the information gap between buyers and sellers in used-good markets. This brings to the fore several questions relevant to markets where consumers have traditionally relied on signals from centralized sources to mitigate purchase frictions. Of particular interest to firms is whether consumers' increased use of decentralized online information *substitutes* or *complements* traditional mechanisms such as certification. Specifically, we examine how the access to online information alters the salience and value of certification for consumers. For instance, with greater online information, are consumers more likely to purchase the quality signal or certification? Do consumers who obtain certain types of online information pay higher or lower prices than others for their purchases?

The used-vehicle market provides the context for our study. Given the complexity of the offerings and the difficulty in determining quality, certification in particular has played a valuable role in reducing frictions in the market for used cars. In recent times, the Internet and the emergence of auto-retailing websites have however dramatically increased the amount of information available to consumers seeking to purchase used cars. This changing

landscape of used vehicle markets makes it an ideal setting to understand the impact of online information on the value of certification – an issue of interest to academicians as well as practitioners. We draw upon a unique and extensive dataset of consumers who report on their acquisition of different types of online information used in their recent used vehicle purchase. The availability of rich accounts of consumers' information search along with transaction details allows us to examine the impact of online information on consumers' choice of certified used cars, as well as the price paid. We develop a simple model motivated by theory and empirical observations from economics and behavioral decision research to explain the impacts of online information. We compare the outcomes of sales where consumers purchased certified used cars with sales of non-certified used-cars, after controlling for a number of buyer, vehicle, and market characteristics. We find that four different types of online information - *comparative product* information, *comparative price* information, *vehicle- specific product* information, and *transaction-specific price* information - have significant but varied impacts on consumers' *value for certification*. Our results highlight the important role of online information for buyer and seller outcomes in markets for used goods. Based on their impact on the demand and the price paid for certified as well as uncertified cars, we find that both specific and comparative price information *complements* certification, while specific and comparative product-related information *substitute* certification. As highlighted later, these findings have significant implications for manufacturers and retailers seeking to leverage the growing power of the Web as well as for third-party information providers.

The rest of the paper is organized as follows. §2 describes the context of our study as well as the increasing importance of online information in the market for used cars. We then

discuss the theoretical underpinnings of the study and offer hypotheses on the impacts of four types of online information on the value of certification in §3. We present details of the empirical study including the data, measures and model in §4, followed by our results in §5. §6 concludes with a discussion of the relevance of our findings for buyers, sellers, and online infomediaries in secondary markets, and suggestions for future work.

3.2. Research Context and Related Works

The used car market is a large and significant one, and has been growing at a phenomenal pace. While 16.5 million new vehicles were purchased in North America in 2006, the corresponding numbers for used vehicles was 44 million (Manheim 2007). In this classic "lemons" market (Akerlof 1970), sellers use different mechanisms to signal the quality of their products.

3.2.1 Certification

The most popular of these quality signaling mechanisms is "certification", which emerged as a byproduct of leasing in the late 1980s and 1990s when luxury car manufacturers and dealers sought to resell vehicles whose lease periods had ended. *Certification* implies that the certified vehicle has been put through a comprehensive inspection process[15]. These certified pre-owned (CPO) vehicles have increasingly become an important category of vehicle purchases. J. D. Power and Associates (2006) estimates that the sale of certified cars (1.6 million in 2006) had increased 46% since 2000, and accounts for over 40% of all used car sales. However, an interesting and crucial aspect of vehicle certification is that, unlike situations where certification is generally provided by independent

---

[15] National Automobile Dealers' Association classifies certification inspections as: general evaluation; under-hood evaluation; exterior assessment; interior evaluation; required service and maintenance assessment; and exterior detailing.

third-parties, guidelines for used car certification is usually specified by the manufacturer, but ultimately provided by the dealers themselves. Certified used cars typically sell for a premium over their uncertified counterparts. Yet, despite the growing popularity of vehicle certification programs, its critics have called into question its benefits relative to the premium (Cutler 2005). Our interest in this study is to examine the impact that the access to decentralized online information has on consumers' *value for certification,* measured along two outcomes - *demand for certified cars* (demand effect) and the *price paid for certified vs. non-certified cars* (price effect).

A well-established stream of analytical research examines certification (e.g., Albano and Lizzeri 2001; Lizzeri 1999, Viscusi 1978), and its role as an effective mechanism to supply quality information in Akerlof-type settings. Empirical studies, while limited, point to the potential value of certification for consumers and firms in markets with information asymmetry such as those for collectibles, antiques, secondary goods, organic foods, and other hard-to-value products (e.g., Dewan and Hsu 2004, Jin and Kato 2006, Terlaak and King 2006, Wimmer and Chezum 2003) [16]. Certification potentially generates new information for all market players. On the one hand, certification plays an allocative role by allowing buyers to choose the type of vehicle that gives them the highest value, thereby, increasing demand from buyers who otherwise may have not entered the used market. On the other hand, buyer sorting also benefits sellers by providing information on buyer's unobservable characteristics such as risk aversion (Ippolito and Mathios 1990, Jin et al. 2010).

---

[16] Some existing empirical works find mixed outcome effects of mechanisms established to signal quality levels for services (e.g., occupational licensing (Kleiner and Kudrle 2000), professional certification (Angrist and Guryan 2008), and mandatory automobile certification (Pratt and Hoffer (1985)).

3.2.2 Online Information

In recent years used car buyers have additionally been able to access information from a variety of online sources that span the gamut from dealer and manufacturer websites, and third party auto review sites (edmunds.com, intellichoice.com, nadaguides.com, vehix.com), to consumer reviews catalogued in user generated sites (autoblog.com, technorati.com). Specialized auto sites offer advice on aspects such as reliability and safety (autocheck.com, carfax.com), information on financing and auto loan rates (bankrate.com, capitaloneautofinancing.com, USAA.com), and pricing specific to vehicle make-model-condition (kbb.com), for instance. Further, online consumers can avail of comparative shopping, and assess differences across vehicle models, based on a variety of attributes - a process that is painstaking and difficult to perform in the offline channel. In keeping with these changes brought about by the Web, consumers are increasingly complementing their personal information search in offline channels with the decentralized information provided by these diverse online sources. The growing popularity of online information is also demonstrated by the fact that since 2004, a greater proportion of used vehicle buyers have found their car through the Internet than both newspaper and magazine classified ads combined (J. D. Power and Associates 2006). Given consumers' growing reliance on online sources it is vital to understand the impact that online information has on the value consumers place on traditional mechanisms used to lower uncertainty, such as certification.

Dimensions of Online Information: We define four different types of information that consumers are likely to seek and obtain online. While used cars are generally less expensive compared to similar model new cars, they are typically of lower expected

quality. Certified used cars, touted by car manufacturers as providing "the reliability of new and the affordability of used", fall in between these two extremes. This price-quality trade-off lies at the heart of consumers' choices. Hence, two distinct dimensions of information become salient in the context of used good purchases.

The first dimension distinguishes between *price* vs. *product* related information about the alternatives (c.f. Kuruzovich et al. 2008). Product and price information play an important role in driving perceived value, which is a function of (perceived) quality and (perceived) price (Zeithaml 1988).The second salient dimension considers whether the online information is specific to the focal used vehicle (a particular used Lexus ES 350) or if it describes characteristics of vehicles of a make-model across its lifecycle (e.g., information about used, certified and new Lexus ES 350). This distinction, which we refer to as *specific* vs. *comparative* information, is especially important in a used good market. Given the extensive uncertainties in used good markets, comparative information may serve as critical reference points in helping consumers form inferences about the price and quality of the used alternatives (Wetzel and Hoffer 1982; Porter and Sattler 1999).

We cross-map these two dimensions (*price-product* and *specific-comparative)* to four categories of information that capture the multi-faceted structure of online information relevant to the purchase of used goods- namely, vehicle-*specific product* information*, comparative product* information, transaction-*specific price* information*, and comparative price* information. These are discussed in further detail in 3.3.

Our study adds to the recent literature that examines the role and impacts of online information on purchase-related decisions across channels. For example, researchers have

analyzed the impact of online information on consumer outcomes such as offline purchase (Alba et al. 1997), price paid (Zettelmeyer et al. 2006) and channel choice (Kuruzovich et al. 2008). Directly relevant to our study are works by Klein and Ford (2003) and Ratchford et al. (2003, 2007), that examine whether online information affects consumers' use of traditional information sources in the context of new automobile purchases. Their findings suggest that the Internet substitutes for time spent with a dealer and with third-party print sources such as ConsumerReports and Edmunds but does not decrease consumers' need for personal sources (friends and relatives). That the use of the Internet may affect consumers' reliance on traditional sources is an interesting and provocative observation – and is also one that motivates our study.

Further, while much of the existing work on online information focuses on new good markets, online information search takes on added significance in the context of used good purchases. However, little is known about consumers' choice and decision-making in used good markets. Also understudied is how online information affects consumers' price outcomes in a market where final price is negotiated upon. Our setting - a large and economically significant secondary market - enables us to investigate the impact of online information on consumers' choices and prices paid for used goods.

3.3 Model and Hypotheses

The quality of a used car is only imperfectly ascertainable before purchase. While consumers may know the average or expected quality of used vehicles of a certain make-model from past experience of self or others and from marketing activities, they often do not know the true quality of any particular used car (Akerlof 1970). They may however employ several cues available in the market to infer quality. In addition to such *quality*

uncertainty consumers face a corresponding *value uncertainty* stemming from not knowing the appropriate price to pay for the vehicle. Thus there is uncertainty about 'what the consumer gets for what she gives' (Zeithaml 1988, p.13). Given our research interest in this paper, we focus on consumers' choice between a certified and a non-certified used car, which are otherwise similar (make-model-year-miles) but differing in expected quality[17]. Faced with quality uncertainty, on the one side, and a potential premium for certification, on the other, consumers choose based on their rational beliefs about seller behavior related to provision of certification and vehicle quality.

3.3.1 Expected Quality of Certified and Non-certified cars in Equilibrium

In the used car market, sellers have traditionally used certification to inform consumers about the underlying quality of the individual product (Lizzeri 1999; Pratt and Hoffer 1986). The presence of certification signals that the quality of the used vehicle lies at or above a threshold q^*_{min} or minimum quality level[18] (enumerated in the vehicle certification checklist). Prior literature on the voluntary disclosure of firms' private information shows that non-disclosure cannot be a pooling equilibrium when sellers of higher quality goods have an incentive to defect- that is, they benefit by signaling their quality (Milgrom 1981; Grossman 1981). At equilibrium, truthful unraveling or "unfolding" occurs from the top until the cost of disclosing exceeds the benefits to the seller. This cost includes the expenses required to raise the quality of the used vehicle up to threshold level specified for certified vehicles. In addition, sellers must take into account or internalize the expected costs of repair for certified vehicles that breakdown or

---

[17] As explained later, in our analyses we control for vehicle brand, year, miles, attributes and options using the VIN.
[18] Leland (1981) shows how the disclosure model is related to a signaling model with implicit costs.

suffer from problems. Thus, sellers (who wish to stay in business) will not find it in their best interest to sell as certified low quality vehicles whose expected costs of post-purchase repair exceed the market value of certification. In other words, certification can be a credible signal resulting in a separating equilibrium if sellers of low and high quality vehicles indeed differ in their expected benefits.

For this result to occur consumers must have common knowledge about the existence of such quality information in the market and sellers must benefit from disclosing private information. Trade literature suggests that an increasing number of car shoppers today are aware of and informed about certification programs (Mitchell 2008), and pay a premium to obtain certification suggesting that it is profitable for sellers. Thus, we expect that high quality sellers will be more likely to reveal their quality using certification. This result has several implications for the distribution of used cars available for sale in the market (c.f. Milgrom 2008).

First, if quality is ex-post verifiable by buyers (e.g., low quality is correlated with breakdown and non-satisfactory performance) and cheating is costly for sellers (high expected costs of repairs for certified cars that default), low quality vehicles are more likely to be sold as non-certified (a notion that finds support in works by Jovanovic 1982 and Lewis 2009). Second, since sellers typically face certification costs that are increasing in the quality difference between threshold and existing quality, used cars whose costs to certify (i.e. raise the vehicle quality to threshold) are much higher than the average certification premium in the market are more likely to be sold as non-certified (Jovanovic 1982). Third, since certification programs are standardized, thus disallowing differential signaling among certified cars of a given make-model, sellers are unable to

credibly convey quality higher than q^*_{min} through disclosure (after controlling for age and mileage effects). As a result of the downward pressure on certified prices, sellers of very high quality certified cars will be worse off if they trade (Ronnen 1991). Very high quality cars may thus not be offered in the used car (certified) market. The available certified cars will tend to be at or close to threshold quality q^*_{min} (Albano and Lizzeri 2001; Milgrom 2008), and non-certified cars will have an expected value lower than q^*_{min} and a much larger quality variance. In the used car market, this reflects the important role of certification in providing consumers with information about the condition of the vehicle, including inspections/repairs that were performed to ensure a minimum standard of quality as specified in a manufacturer checklist.

The above implications closely resemble the outcomes observed in real world used-car markets. Over the years, as certification programs have matured and a greater share of consumers are informed and aware of such programs, the average quality of (certified) cars that are traded has also improved. Sultan (2010) finds that non-certified cars required more maintenance expenditures than certified cars of a similar age. Furthermore, non-certified cars present a higher purchase risk given their higher quality variance, even if they do not differ much in expected quality levels. Stolyarov (2002) provides evidence of a double-hump regularity in used auto sales which supports our abovementioned arguments on the equilibrium distribution of used cars in the market. There are high sales of vehicles 3-5 years old and about 10 years old. Young vehicles less than 3 years have very low re-sale rates suggesting that owners of such high quality used vehicles prefer to keep owning them (or wait to sell) rather than obtain a price corresponding to threshold quality in the certified used market. Vehicles in the middle

age group are unlikely to fetch certified prices, and are also perhaps valued higher than the price at which non-certified used cars sell. Their owners thus benefit from continuing to use the cars rather than sell them at lower average prices, leading to the observed patterns.

Before we examine the impacts of online information, as a baseline, we first establish that consumers do indeed perceive differences between the expected quality of certified and non-certified cars. When consumers believe or conclude that certified vehicles are on average higher quality and lower variance than non-certified cars, this should be reflected in differences in their willingness to pay for the two types of used vehicles[19]. Our first hypothesis states that,

> HYPOTHESIS 1a. *Buyers pay a higher price for a certified car compared to a similar non-certified one.*

Next, we examine how the availability of online information alters consumer's valuation of certification and impacts the demand and WTP for certified vs. non-certified used cars.

3.3.2 Role of Online Product Information

We summarize our main arguments for product information here before we examine each type of information separately. Past work in marketing literature has found that improving a (competitive) disadvantage attracts consumers from alternatives more than does improving a (competitive) advantage (Heath et al. 2000). When a price-dominant alternative reduces its disadvantage in quality, its ability to attract demand

---

[19] It is possible that consumers are willing to pay more for certification due to the availability of a warranty. We control for warranty in our analyses. Support for H1 then implies that consumers value the quality signal.

away from a competing alternative is particularly significant since it offers more (perceived) quality at the same lower price. The implication for our work is that learning (which reduces uncertainty) about a product and its quality raises consumers' perceived value more for non-certified cars than certified cars.

Conditional upon choosing a certified or non-certified car, online product information can also impact consumers' WTP when prices are negotiated. As a first order effect, learning more about the product and its features allows consumers to lower their uncertainties and find products that fit their preferences better. This raises their WTP. Additionally, online product information has a second-order effect on WTP that arises due to the important role that relative quality or quality differentiation plays in markets where quality uncertainty is a prevailing factor and absolute quality is difficult to ascertain. In the context of services, Boulding et al. (1993, 1999) and Inman et al (1997) propose that competitive alternatives enter explicitly into consumers' evaluation of the focal service. In particular, Boulding et al. (1999) find that holding fixed the level of the focal service, a higher level of quality associated with a competitive alternative decreased the evaluated quality level of the focal service. Similarly, product information about the quality of used goods may alter the level of quality differentiation that consumers perceive. Certified cars present a unique proposition - higher quality at a premium. Consequently, consumers that buy certification will value it and pay more when the quality of certified car is sufficiently differentiated from the non-certified car. Conversely, reduction in the quality differentiation raises a non-certified buyers' WTP.

We show below that specific and comparative product information provide consumers with different types of knowledge about used vehicles, and therefore impact consumers' demand and WTP differently.

Product Information on Features and Specifications: Vehicle-specific product information found through online sources provides knowledge about a particular used vehicle. The Internet makes it easier for consumers to efficiently search large and complicated product spaces, and thereby plays an important role in allowing consumers to locate and learn about their particular used car. Vehicle-specific product information offers details on the various features and options available (e.g., airbags, ABS brakes, anti-theft locks, parking aids), the external and internal conditions combined with photographs and descriptions of the specific vehicle. This information may be broadly labeled as "search" attributes, referring to the fact that these characteristics are observable to the consumer with certainty upon pre-purchase inspection. Yet, to an untrained eye or inexperienced car buyer, the sheer number and variety of options available on cars today renders it difficult to learn about these features from merely visiting the dealer. Online sources may therefore aid consumers in becoming better informed about such search features[20].

Prior studies have found that when a product consists of a high proportion of difficult-to-assess experience attributes, consumers may infer unobservable quality from observed product features either basing their inferences on certain correlated attributes or overall evaluations (Dick et al. 1990). In the used car market, customers who exhibit such tendencies will associate better fit of vehicle attributes with higher levels of unobserved quality. However, consumers that have alternate means of assessing quality will be less likely to

---

[20] The availability of features is already reflected in the seller's asking price; we are interested in examining whether learning about these search attributes may have additional impacts on the potential value provided by certification.

make inferences from proxy attributes (Zeithaml 1988). Such an inference effect will therefore be more pronounced for buyers of non-certified vehicles with higher quality variance; whereas the presence of a quality signal in certified cars reduces consumers' likelihood of generating quality inferences from search attributes[21]. For instance, learning about the features on a well appointed non-certified used car may lead consumers to judge it as being of higher quality than in the absence of such information. As a result, consumers that make such inferences will lower their valuation of certification, and be less likely to purchase certified cars. Thus, with vehicle-specific online information about product search attributes, more consumers will prefer non-certified cars, thereby increasing its demand.

HYPOTHESIS 2a. *Vehicle-specific product information obtained from online sources reduces buyers' likelihood of purchasing a certified car.*

Vehicle-specific information on available product features may also influence buyers' willingness to pay for certified vs. non-certified cars. Past work has shown that learning about product features benefits heterogeneous consumers by allowing them to find better fitting products and reducing their price sensitivity, leading them to be willing to pay more (Boulding et al. 1994; Kaul and Wittink 1995; Mitra and Lynch 1995). Since information on search attributes is more likely to bear new information for non-certified cars, buyers may pay higher prices than in the absence of vehicle-specific product information. Further, improvement in the perceived quality of non-certified vehicles also lowers the perceived quality differentiation between certified and non-certified vehicles, which additionally raises consumers' WTP for non-certified cars.

---

[21] Search attributes by themselves need not be correlated with (unobservable) true vehicle quality and reliability. Further, certification checklists provide detailed information on the availability of features and options, thereby reducing the value of such information obtained from online sources.

In comparison, consumers who buy certified vehicles do not obtain new information from learning about product search attributes, and therefore do not make similar inferences about vehicle quality. Since they are unlikely to perceive a lowering in the expected quality differentiation, we conjecture that the WTP of certified car buyers is not systematically affected by the availability or lack of vehicle-specific product information.

HYPOTHESIS 2b. *Vehicle-specific product information increases the price paid by buyers for non-certified cars; while it is likely to have no significant effect on the price paid for certified cars.*

Product Information on Vehicle Quality: A second type of product information provides consumers knowledge about experience or quality attributes- which are, in contrast to search attributes, harder to ascertain prior to using the vehicle. Several websites on the Internet specialize in providing model reviews, consumer and expert ratings, and results for test drive, handling and crash-outcomes for automobiles. This extensive information about vehicle reliability and safety for new vehicles makes it feasible for consumers to gather knowledge about the quality of the focal used vehicle when it was new. Comparative product quality information pertains to the class of all vehicles of a particular make-model-year and not to any one specific vehicle; however, it may serve as a useful reference point for consumers to infer the residual quality of their focal used vehicle. Such effects have been observed by Sullivan (1998) and Purohit (1992) in the automobile market, and by Janakiraman et al. (2009) in other settings. They find evidence that when faced with uncertainty, consumers' perceptions of quality of known products spill over onto other products within the same brand about which less is known. As a corollary, learning about a

(new) 2006 Toyota Camry LE's performance related to engine, transmission, driveline, steering and suspension, can help mitigate consumers' performance uncertainty associated with purchasing a used 2006 Toyota Camry LE. Moreover, several used vehicle advertisements also often highlight the tagline that "The best new cars make the best used cars", referring to the notion that higher quality new vehicles are also likely to retain more quality when sold as used than models that start out lower in quality.

Since consumers face high levels of quality uncertainty in this market, comparative product information about the unobserved quality of vehicles is important for both certified and non-certified used vehicles. However, such quality information is more valuable for non-certified cars because of their greater variance in expected quality. Whereas for certified cars, this perceived quality increase may be limited because comparative product quality information provides less new information for certified vehicles, both due to the lower variance in quality and since some of the online quality information may even be redundant (also made available as part of the certification checklist). We therefore suggest that the availability of comparative product quality information helps increase the demand for non-certified vehicles by making it attractive to some consumers who would have otherwise preferred to buy certification in the absence of such information

HYPOTHESIS 3a. *Comparative product information obtained from online sources reduces buyer's likelihood of purchasing a certified vehicle.*

Next, we examine the effect on WTP. Comparative product information benefits the non-certified vehicle in two ways. First, the improvement in expected quality (and variance) increases buyers' WTP for non-certified vehicles than in its absence. Second, the decrease in perceived quality differentiation between certified and non-certified alternatives also raises

consumers' valuation and WTP for non-certified cars. For certified cars, these two effects act in opposing directions. On the one hand, information that improves perceived quality lends further credibility to the certification signal and may increase the value of certification for the buyer. On the other hand, the lowering in perceived quality differentiation lays a downward pressure on consumers' valuation of certified cars, similar to the ironical observation about a popular restaurant made by Yogi Berra, that "Nobody goes there anymore; it's too crowded.." In our context, when consumers perceive the used vehicle to be of sufficiently good quality, they may lower their valuation for certification. The net effect on WTP for certified cars depends on the relative strengths of the counteracting effects. We hypothesize that,

> HYPOTHESIS 3b. *Comparative online product information increases the price paid by buyers for non-certified cars; while it is likely to have no significant effect on the price paid for certified cars.*

3.3.3 Role of Online Price Information

The unique features of pricing in the used car market (and more generally, the car industry) – including that final prices are negotiated, prices vary much and often across dealers, and consist of multiple components- suggest that price information may play an important role in buyers' purchase outcomes. A recent study by Busse, Simester and Zettelmeyer (2009), finds that car buyers are influenced not only by actual price information but also price cues. Not surprisingly then, the Internet has spawned numerous websites that provide valuable price-related information to consumers in the auto market

Online channels allow consumers to research and discover better financing rates, and become aware of available incentives and offers. Such information is likely to lower consumers' perceptions of the price of owning a used vehicle. Following from the notion that

reducing a (competitive) disadvantage is more valuable than improving a (competitive) advantage (Heath et al. 2000) we expect that improvements to the perceived price of alternatives will asymmetrically attract more demand towards the higher-priced certified cars than non-certified cars. This is supported in past work that documented that when prices are lowered, more consumers switch up to higher quality (and price) national brands from store brands rather than vice versa (Blattberg and Wisniewski 1989; Sivakumar and Raj 1997). The recent cash-for-clunkers program, which gave buyers a credit of up to $4500 towards the purchase of a new (and more efficient) car, also remarkably increased sales of the new vehicles, many from buyers who would have typically spent less and bought used cars otherwise (Bunkley 2009). The result of this effect that lowers the perceived price of owning a car is to incentivize consumers to spend more as a result.

A second effect of learning about prices in the market is to help buyers locate the seller with the lowest price and learn about the distribution of average prices in the market across new and used alternatives - information that can be useful in the bargaining process. In past work, it has been documented that consumers who obtain relevant price-related information from online sources use it to negotiate with the dealers for lower prices on their vehicle (Busse et al. 2006; Zettelmeyer et al. 2006). In recent experimental work, Mazar et al. (2009) demonstrate that price is a powerful contextual variable – and affected participants' willingness to pay as they made inferences from the price distribution—going down for the left skewed distributions and up for the right skewed distributions. A key effect of learning about (low) price information then is to lower consumers' WTP. Whether consumers pay a different amount upon obtaining price-related information than without depends on the

relative direction and strengths of these two effects. We discuss each type of price information next.

Price Information on Financing and Incentives : The final price of a used car consists of several components. On the base price of the vehicle are included the price of additional "car" add-ons such as certification/warranties less any available incentives or promotions. Further, when the vehicle is financed, consumers also care about the interest rates and the resulting monthly payments. The Internet makes it relatively easy for buyers to access information related to financing rates and incentives associated with purchasing a used vehicle. Obtaining lower price information will in turn create favorable price perceptions in the consumer's mind. This information denotes rates and offers that are applicable to the specific transaction for the car chosen by the consumer-hence, we refer to it as vehicle or transaction-specific price information.

A typical concern of used car buyers is that certification is expensive. A quick observation of the auto market suggests that dealers are more likely to offer special rate financing schemes and more incentives for certified cars, mimicking the low rates available on new cars. Obtaining such information aids buyers to reduce monthly payments and can lower the perceived price of certified cars. We anticipate that reductions in the perceived price will increase the attractiveness and demand for the more expensive certified cars vis-à-vis non-certified cars for reasons outlined below.

Research on mental budgeting suggests that when consumers have budgeted an amount to a purchase - here, a used car - (unexpected) favorable changes in the price of the product may produce a perception of savings from the transaction and may result in a congruent spillover effect (Thaler 1999). In other words, such perceived price savings may

lead consumers to re-invest the surplus to buy more product (features/quality) for the money- a claim that has been supported in several empirical works (e.g., Ariely et al.2003; Heath and Soll 1996; Heilman et al. 2002; Janakiraman et al. 2006). Such a "savings effect" therefore encourages consumers to spend more leading to increased demand for high quality/high price alternatives. This is supported by recent findings from the auto industry that customers who were offered incentives (e.g., employee discount programs) thought that the overall prices were good enough that they could afford to buy a more expensive car (Busse, Simester and Zettelmeyer 2009). In related work, Gourville (1998) provides a related explanation that information about monthly payments induces a favorable temporal reframing that shifts consumers' attention to smaller, less aggregate, ongoing expenses from larger, more aggregated and one-time expenses, thereby, reducing consumers' perceived transaction costs. Transaction-related price information therefore plays an important role by increasing the attractiveness of and demand for the certified used vehicle. We accordingly posit that,

HYPOTHESIS 4a. *Vehicle/Transaction specific price information obtained from online sources increases buyer's likelihood of purchasing a certified vehicle.*

Conditional on choosing a used vehicle, the main role of price is as a measure of economic sacrifice. Thus, a direct effect of price-related information is to aid consumers in negotiating and paying lower prices for their used vehicle. The effect on WTP for non-certified cars is to lower the WTP or the price paid by the buyers. The WTP for certified car buyers, in turn, is determined by two effects. First, the knowledge of price information enables them to bargain more successfully and obtain lower prices. Second, the increased attractiveness of certification due to perceived lowering of its cost may influence consumers to spend more. We expect that the net effect of these two drivers will result in price paid by

consumers for certified cars to not be significantly different in the presence of transaction-related price information than in its absence. We hypothesize that

HYPOTHESIS 4b. *Vehicle/Transaction specific price information lowers the price paid for non-certified cars; while it is likely to have no significant effect on the price paid for certified cars.*

Price Information on New, Used, and Trade-Ins: The second kind of price information allows consumers to learn about the market prices across a portfolio of available alternatives for a given make-model. For a vehicle of interest (say Audi A6), consumers may learn about the current average market value of the used good as a trade-in, as a retail offering across different quality conditions, and also as a new vehicle. Further, consumers may also learn about dealers' invoice prices. Such information is comparative because it does not reflect the true (asking) price of any particular used vehicle; but rather serves as a reference point that the consumer may use to derive fair price estimates for their particular used vehicle. Learning about prices ranging from the low-end trade-in to the high-end new vehicle may more importantly invoke contextual inferences in a consumer who is purchasing a used vehicle. In the presence of uncertainty about the quality and fair price to pay for the used good, consumers may look to information available from the context (here, prices of new and used) to generate such inferences (Kamenica 2008). In particular, consumers have been known to demonstrate a compromise effect (Kivetz et al. 2004; Simonson 1989; Wernerfelt 1995) or Goldilocks prici 0png effect (Shapiro and Varian 1999), whereby the inclusion of extreme-priced alternatives has been observed to increase the demand for the middle option. Such an effect arises when consumers tradeoff the higher price but also higher

expected quality of new vehicles with the lower price but also lower expected quality of non-certified used vehicles. In doing so, the middle option is often deemed to be the most attractive. In our context, this would be the certified car. We therefore expect that,

HYPOTHESIS 5a. *Comparative price information obtained from online sources increases buyer's likelihood of purchasing a certified vehicle.*

Once again, obtaining price-related information allows consumers to locate and bargain for lower prices on their used vehicle purchase. This impact should affect both certified and non-certified cars alike conditional on choice. In addition to this effect, for certified car buyers, there is a positive effect on WTP arising from the increased attractiveness of certified cars as middle option. Information about high-priced new cars enables consumers to better appreciate the value (price-quality tradeoff) of almost-new certified used vehicles, and increases their willing to pay for it. This is consistent with earlier marketing research on the compromise effect which finds that the presence of an extremely high-priced product alternative can increase the willingness-to-pay for more moderately priced products within a product category (e.g., Krishna et al. 2006). In our context, information that increases the value for certified cars also reduces the value for non-certified cars in a choice between the two. These effects of lowered WTP from better and informed bargaining and higher WTP due to greater attractiveness may counteract each other and result in no change in consumers' WTP for certified cars with and without comparative price information. In contrast, as a consequence of comparisons across new and used alternatives, buyers who purchase the non-certified used vehicles may generate unfavorable quality inferences from comparative price information. If it were indeed the case, such an inference would lead buyers to perceive a greater quality differentiation between certified and non-

certified vehicles, which then additionally lowers their WTP. Thus the ability to negotiate lower prices is reinforced by the lower perceived attractiveness of the used car for buyers of non-certified cars. Thus, we expect that

> HYPOTHESIS 5b. *Comparative price information lowers the price paid for non-certified cars; while it is likely to have no significant effect on the price paid for certified cars.*

3.4 Empirical Study

3.4.1 Data

Our study is based on secondary data obtained from a survey of buyers, who purchased 1999 to 2004 model year used vehicles in late 2003, conducted by one of the largest market-research firms. The quota sampling strategy was designed to ensure that a sufficient sample size was obtained for car-make analysis, ensuring a minimum return of 125 purchases for nameplates with certification programs (and 120 for others). Two versions of an eight-page questionnaire along with a $1 incentive were sent out in late January 2004 (within 3 months of purchase), followed by a reminder postcard after a week. Out of the total mail-out to a randomized sample of 78,534 buyers, 12,142 surveys were returned resulting in a response rate of 15.5%. The dataset consists of both consumers who used the Internet as part of their purchase process and traditional consumers who did not use the Internet. Sampling weights are used to ensure that the distribution of makes in the sample was representative of the total personal use registrations of vehicles completed during the sampling period.

We follow Zettelmeyer et al. (2006) in defining a "car" as the "interaction of make, model, body type, transmission, displacement, doors, cylinders, and trim level" (p.170). A "car" is measured using the first 8-digits of the vehicle identification number (VIN), and

allows us to adequately control for vehicle fixed effects. We restrict our analysis to the top

135 "cars", each with at least 25 vehicle purchases, resulting in a total of 5,349 observations.

Given our interest in comparing the effects of online information on certified and non-

certified purchases, we only retain "cars" with both types of sales[22]. Finally, we restrict our

sample to cars purchased at dealerships, resulting in 126 "cars" with 3213 purchases, of

which 35% were certified vehicles.

3.4.2 Measures

Table 3.1 summarizes the variables used in our empirical analyses. Our primary outcome

variables are the choice of vehicle (certified or non-certified) and the price paid for the used

vehicle. $CERTIFICATION$ is a binary measure, while $PRICE$ is measured in dollars. The

independent variables are measured as follows. Online $INFORMATION$ is categorized into

vehicle/transaction specific product ($SPROD$) and price ($SPRICE$) information, and

comparative product ($CPROD$) and price ($CPRICE$) information. Buyers report on a

multitude of online information found (0/1) by them during the course of shopping for the

used vehicle they purchased (see Table 3.2).

We employ principal components factor analysis with varimax rotation to reduce the

dimensionality of online information into a set of four meaningful factors (with eigenvalues

greater than one) that explains 79.98% common variance[23]. $SPROD$ includes access to

vehicle photographs, and tools for assessing available features and specifications of the

specific used vehicles. $SPRICE$ includes information on special offers, discounts and

---

[22] We dropped 10 "cars" (68 luxury and 419 non-luxury purchases), resulting in a marginal decrease of certified purchases (33.09% vs. 33.78%), and slight increase in price paid. The new sample did not differ significantly in online information found by consumers.

[23] We use factor analysis as means of reducing dimensionality while extracting underlying patterns across items. Factors exhibit acceptable reliability (0.66 to 0.81), with main factor loadings >0.62 and cross-loadings < 0.35.

financing options, along with warranties and certification programs available on specific used vehicles of interest to a potential buyer. $CPROD$ includes historical performance, reviews, safety and crash test ratings, and road-handling abilities of vehicles in a given make-model class. $CPRICE$ includes prices of new vehicle alternatives and trade-in values for vehicles in a given make-model class.

The dataset includes $VEHICLE$ characteristics such as $Mileage$, $Model\ year$, and $Luxury$ make. We construct 126 dummies to represent the "cars" described earlier. The dataset also contains information on $BUYER$ demographics such as $Gender$, Age, $Income$, $Low\ Education$, $Minority$ and $Marital\ Status$. Buyer's pre-purchase intention is captured using dummies $Want\ Used$, $Want\ New$, or $Open$ - to indicate the buyer originally only wanted to purchase a used vehicle, new vehicle or was open to both used/new. $Previous\ Car\ Used$ is equal to 1 if the buyer previously owned a used car. We include several additional controls in our analyses.

Buyer's access and use of offline information to search and locate the seller and/or purchased vehicle is captured by their use of classifieds/ads in newspapers, TV and magazines ($Offline\ Ads$), and recommendations from friends, relatives and/or own personal experience ($Offline\ Personal$). We use dummies to indicate whether the buyer obtained vehicle $History\ Reports$ prior to purchase (mostly provided by dealers in our sample), and whether buyers $Visit\ Dealer\ After\ Online\ Search$. Market characteristics include dummies for $Metro$, $Small$, or $Rural$ indicating the type of market where the car was purchased. Seller type indicates whether the car was purchased from a $New\ Car\ Dealer$ or $Used\ Car\ Dealer$. Two measures –buyer's overall $Satisfaction$ with the vehicle (1-10 scale), and number of $Post-purchase\ Defects$ encountered within three months of

vehicle purchase – act as a proxy for unobserved short-term *Vehicle Quality*. Finally,

Manufacturer warranty and *Add warranty* are binaries equal to 1 if the vehicle had

remaining manufacturer warranty and if the buyer purchased additional warranty (extended

service contracts from the dealer)[24].

3.4.3 Empirical Model

Our primary dependent variables are log price in equation (1) and the choice of certification

in equation (2), which are modeled as a simultaneously affecting each other. In addition,

following prior work that finds that consumer's decision to obtain and use online information

in automobile purchases is affected by several contextual and personal factors (Kuruzovich et

al. 2008, Ratchford et al. 2003, 2007) we model consumers' acquisition of online information

as endogenous in equation [3]. Below, Information = {CPROD, CPRICE, SPROD,

SPRICE}.

$PRICE =$

$\alpha_0 + \alpha_1 VEHICLE + \alpha_2 BUYER + \alpha_3 CERTIFICATION + \alpha_4 INFORMATION +$

$\alpha_5 INTERACTIONS + \alpha_6 CONTROLS_P + \alpha_7 Z_{price} + \varepsilon_{Price}$

$$(1)$$

$CERTIFICATION = \beta_0 + \beta_1 VEHICLE + \beta_2 BUYER + \beta_3 PRICE + \beta_4 INFORMATION +$

$\beta_5 INTERACTIONS + \beta_6 CONTROLS\_C + \beta_7 Z_{certification} + \varepsilon_{certification}$

$$(2)$$

$INFORMATION = \gamma_0 + \gamma_1 VEHICLE + \gamma_2 BUYER + \gamma_3 CONTROLS\_I + \gamma_4 Z_{information} +$

$\varepsilon_{information}$ $\quad (3)$

---

[24] Additional warranty refers to bumper to bumper or powertrain warranty purchased but not included with certification.

The $INTERACTIONS$ vector is a set of centered cross products between

$CERTIFICATION$ and the $INFORMATION$ factors, and two interactions of

CERTIFICATION with $Mileage$ and $Luxury$ make, to capture the differential impacts of

certification across high mileage and luxury cars. Equations [1] - [3] include a vector of

common $VEHICLE$ and $BUYER$ variables, and offline information search variables-

$Offline\ Ads,\ Offline\ Personal,$ and $Visit\ Dealer\ After\ Online\ Search$ dealer. Controls

common to $CONTROLS\_P, CONTROLS\_C$ and $CONTROLS\_I$ include

$Want\ Used, Want\ New, Open$ to vehicle type, $Vehicle\ Quality,\ Satisfaction,$ market

size, and seller type. Additionally, $CONTROLS\_P$ contains $Post-purchase\ Defects,$

$History\ Reports,\ Manufacturer\ warranty$ and $Add\ warranty.$ Finally $Z_{price},$

$Z_{certification},$ and $Z_{information}$ are vectors of instruments that enable estimation of our

system of simultaneous equations (1) – (3), as discussed below.

3.4.4 Estimation Procedures

While certification is not provided separately to consumers, the availability of detailed

consumer, vehicle, and transaction-related controls as highlighted in Table 3.1 facilitate

sophisticated estimation procedures that enable us to tease apart the impact of information on

outcomes. We address several concerns that arise in estimating the parameters of interest in

our system of equations.

Treatment bias: Non-random selection into treatment conditions (here, the choice of

certification) in the sample leads to biased coefficient estimates if ignored (Heckman 1979).

In our study, treatment bias may arise from either demand-side or supply-side selection

effects caused by unobservables. Demand-side selection arises when unobserved variables

lead buyers to both purchase certification and obtain systematically higher or lower prices.

For instance, risk averse buyers are more likely to buy certified used cars but also likely to pay higher prices on average (if they believe higher prices proxy for higher quality), compared to buyers who are less risk-averse. Alternatively, savvy car buyers may be able to negotiate better prices for their certified vehicle, leading to a negative selection effect. In these cases, the coefficient of *Certification* in equation [1] would be over/under-estimated, respectively, as it captures not only the effect of certification, but also that of the correlated unobservables, on price. We simultaneously estimate the price-certification equations to control for treatment effects using a selection correction term and an exclusion restriction[25]. A likelihood ratio test suggests the absence of demand-side unobserved selection in the purchase of certified vehicles- the correlation between equations [1] and [2] is .07, but the test of independence of equations is not rejected ($\chi^2(1) = 1.27$, $p = .26$).

Similarly, there may be non-randomness in the vehicles that are chosen to be certified by the seller, leading to seller-side selection effects. Sellers may selectively choose to certificate certain types of vehicles (e.g., newer-model, low mileage, and luxury makes) that are more profitable to sell as certified, thereby upwardly biasing the coefficient of certification on price. The coefficient between *Price* and *Certification* in our sample is .19 (*p < .01*), suggesting either, that higher value cars are certified, or, that certified cars are priced higher. We deal with this issue by adding *Price* as an explanatory variable in [2].

Error covariance: Another concern relates to the possibility of contemporaneous error covariance across equations (1) – (3) for a given buyer, indicating that common unobservables influence consumer's information acquisition, choice of certification and price.

---

[25] The model includes a selection correction term that rests on the assumption of bivariate normality of residuals across (1) and (2). We additionally include *Shop Certified,* a dummy, to indicate buyers' pre-purchase intention to specifically shop for a certified vehicle. The use of an exclusion restriction helps by reducing reliance on the functional form of the equations for identification (Greene 2002).

A related issue arises from the presence of a common subset of right hand side regressors that could also potentially confound the parameters of interest.

Endogeneity: OLS assumptions may be violated due to the presence of reciprocal endogeneity between the choice of certification (1) and price (2) equations (ε_{Price} and $\varepsilon_{certification}$ are correlated) as discussed above. Another source of endogeneity in the system arises from omitted variables that affect online information and price (leading ε_{Price} and $\varepsilon_{information}$ to be correlated), and online information certification (leading $\varepsilon_{certification}$ and $\varepsilon_{information}$ to be correlated). Omitted variables that simultaneously affect the likelihood of obtaining online information, and the availability of certification and/or price would confound the coefficients of $CPROD$, $CPRICE$, $SPROD$ and $SPRICE$ in (1) and (2). For example, popular models have a larger number of websites dedicated to them, and a greater availability of certified cars (due to larger volume of leases/trade-ins). Thus, finding online information is potentially endogenously determined by unobservables driving the vehicle's price and certification status. In another example, if consumers who are likely to bargain heavily and pay lower prices are also more likely to obtain online information (because it is more valuable to them), then the effect of information on outcomes will be overestimated.

In order to take into account both the effects of cross-equation *error covariance* and *endogeneity,* we employ the three-stage least squares technique (3SLS) that combines 2SLS and SUR (Greene 2003; Wooldridge 2002)[26]. Certification and online information are endogenous in equation (1), and online information and price are endogenous in equation (2). The set of instrumental variables Z_{price}, $Z_{certification}$, and $Z_{information}$ helps identify our

---

[26]The Hausman test of no endogeneity is rejected, supporting the use of simultaneous equations, while the specification test of 2sls vs. 3sls is not rejected, suggesting that in the absence of model misspecification, 3sls is more efficient.

system of equations (see Table 3.4 notes). The validity of the instruments is assessed by ensuring that the instruments are orthogonal to the residuals of the structural equation in which they serve as an instrument. In equations (1) and (2), the heteroskedasticity-robust overidentification test statistic Hansen-J is not rejected, ensuring validity. Instrument relevance is assessed using a combination of first-stage F-statistic, Shea partial R^2, under-identification and weak identification tests. The details are provided in the Appendix.

3.5 Results

3.5.1 Main analyses

Table 3.3 reports the summary statistics across the certified and non-certified sub-samples. Table 3.4 presents the main results of our 3SLS analyses. A linear probability model such as 3SLS assumes that the dependent variable is continuous but this is not the case for the choice of certification. Aldrich and Nelson (1984)and Angrist and Krueger (2001) demonstrate, however, that this is not necessarily a fatal problem since the consistency of the second-stage estimates does not depend on getting the functional form of the first-stage correct. Therefore linear regression is sufficient for our purposes. More importantly, we find robust results across our estimations using OLS/2SLS and 3SLS, regardless of whether the dependent variable is binary or continuous. We therefore believe that our results indicate a clear relationship between information and purchase outcomes. Next, we discuss our main results from the 3SLS model.

The impacts of *VEHICLE* Characteristics are shown in Panel A; effects of *INFORMATION*, *CERTIFICATION* and *INTERACTIONS* are in Panel B; *BUYER* demographics and psychographics are in Panel C; and Panel D includes offline information

controls. Other control variables described earlier are included but not displayed. We outline our main results starting with the impacts on the retrieval of online information (models M1c—M1f), choice of certification (model M1b), followed by price paid (model M1a).

Acquisition of Online Information: We find that vehicle characteristics- mileage and model year- do not impact buyer's likelihood of obtaining online information. Several buyer characteristics on the other hand were significant. While demographic factors such as income, gender, and marital status did not matter, differences in minority status, education, and age affected what online information buyers obtained – with minorities and less-educated consumers less likely to obtain any online information. Additionally, consumers who previously owned used cars were less likely to seek comparative product information, while those who specifically shopped for a used car obtained more vehicle-specific product details and less information on new car offers and prices. Further, we observe that buyers who used classified/ads in offline channels were more likely to seek information online; while buyers who relied more on their friends/relatives were less likely to do so.

Choice of Certification: We observe that buyers in our sample are more likely to purchase certification on higher mileage, older model, and more expensive cars. Further, being a female, being older, with previous ownership of used cars, and being in the market specifically for used cars, were associated with a higher likelihood of buying certified cars. Interestingly, the use of both offline impersonal and personal sources increased likelihood of purchasing certification. After controlling for the above factors, we find that online product and price information have disparate impacts on buyer's propensity to buy certified vehicles. Both comparative and specific product-related information *reduced* the likelihood of purchasing a certified vehicle. However, the retrieval of comparative and specific price

106

information *increased* the likelihood of a certified purchase. Thus, hypotheses H2a, H3a, H4a and H5a are supported.

Price paid: In the price model (M1a), we find that on average, buyers pay $80 less for each additional 1000 miles and an additional $1,645 more for a car newer by one year. As hypothesized (H1) our results show that buyers pay a premium (13.1% or $2,060 for an average car) for certification over a comparable non-certified vehicle. We however fail to observe any additional premium for certification of luxury and lower mileage cars[27]. On the one hand, luxury vehicles cost more to repair. Thus, in terms of costs of post-purchase repair and maintenance, obtaining a vehicle with lower risk or quality variance is more valuable for luxury cars than for non-luxury cars. Buyers may consequently attach more value to certification on luxury used vehicles. On the other hand, luxury cars are generally touted as having greater expected quality. If used luxury vehicles retain greater residual quality than similar aged non-luxury vehicles, buyers expect non-luxury used cars to display greater variance in quality (in addition to lower expected quality than luxury cars). If consumers perceive greater quality differentiation between certified and non-certified alternatives of non-luxury used cars, it follows that they will then place more value on certification for non-luxury used cars, and pay a higher premium. These effects are possibly masked in our aggregate estimates. The absence of additional premium for low mileage cars is likely an artifact of the mileage restrictions for certified cars.

As for the impact of online information, we find once again that product and price related information have opposite effects on price paid. The main effect coefficients suggest that acquiring comparative and specific product information *increased* the price paid, while

---

[27]J.D. Power and Associates estimate the certification premium to be $1000 for non-luxury cars and $3000 for luxury cars (2003), while the corresponding CNW Marketing Research estimates are in the range of $300 - $1,750 and $2,100 - $3,200 (How Much Does Certification Cost 2008).

comparative and specific price information *lowered* it. Recall that our interest in H2b, H3b, H4b and H5b is to examine whether obtaining online information led buyers to display greater willingness-to-pay for certified cars when compared to non-certified cars. The relevant results are the interaction effects (Table 3.4, panel B) between certification and the four information factors, all of which are insignificant in M1a. This suggests that either, consumers who obtain online information do not pay significantly different prices for certified versus non-certified used cars (due to better bargaining position, for instance) or, that the price paid for a certified car is not influenced by whether the buyer obtained online information or not. In order to tease out the effect of information on each type of used vehicle purchase, we proceed to analyze results for sub-samples of certified (M2a) and non-certified (M2b) cars as reported in Table 3.5.

While the impacts of online information on price paid for certified vehicles remains insignificant, we observe a significant influence of online information on the price paid for non-certified used cars. We interpret the estimated coefficients on price paid as the relative change in price arising from one unit change in the regressors for the mean vehicle. For an average non-certified car in our sample, comparative product information increases price paid by $430 per unit of information, while specific product information increases price paid by $180 per unit of information, representing a premium of 1.2% - 3% of vehicle price. Comparative price information reduces price paid by $260 per unit of information obtained; whereas specific price information lowered price paid by $40 per unit of information, equaling .2% - 1.8% savings for an average used vehicle. Thus, hypotheses H2b-H5b are indirectly supported through the varied effects of online information on price paid for non-certified cars. Considering that consumers obtain multiple pieces of each type of online

information, our estimates of the impact of per unit online information are comparable with previous research that found that price-related information provided consumers with 2% - 16% price savings (Baye et al. 2003, Brown and Goolsbee 2002, Zettelmeyer et al. 2006).

In summary, both types of product information were found to lower consumers' choice of certification. However, as seen from the estimates above, the price paid by non-certified vehicle buyers was more strongly influenced by comparative product information $CPROD$ (than $SPROD$). This is not surprising since in a market such as the one for used cars, information about unobserved vehicle quality and reliability would be more valuable to consumers than merely learning about available features and specifications (controlling for vehicle characteristics). Similarly, obtaining $CPRICE$ induced buyers to pay less for non-certified cars than obtaining $SPRICE$ information. While both types of price information endows buyers with information critical to successful bargaining, if $CPRICE$ additionally leads consumers to infer quality from the prices of new and certified goods, buyers may be willing to pay lower amounts than with only $SPRICE$.

3.5.2 Post-hoc analyses

We attempt here to shed some light on the process by which information affects outcomes. We conduct two sets of post-hoc analyses. In the first, we examine the reasons for purchasing their chosen vehicle provided by consumers who switched from originally intending to buy one type of vehicle (certified or non-certified) but then purchasing another. This subset of buyers altered their purchase decision as a result of obtaining online information, and controlling for other factors in the empirical model, allows us to more cleanly attribute the change to the information obtained. In the survey, buyers report on their purchase related behaviors and reasons why they purchased the used car that they did. First we assess whether

obtaining the two types of product related information $CPROD$ vs. $SPROD$ influenced

consumers in different ways to switch to buying non-certified vehicles. We find that those

who obtained more $SPROD$ than $CPROD$ bought non-certified cars because they found

vehicles that closely fit their needs for vehicle features and specifications. Whereas those

who obtained more $CPROD$ than $SPROD$ were relatively more sensitive to quality related

information and rejected vehicles based upon their history reports (see Table 3.6a). In terms

of price paid, we find that non-certified buyers who believed that certified cars were better

quality than non-certified cars paid *less* for their non-certified cars ($106 less vs. $268 less).

Next, we examine buyers who switched from originally wanting non-certified to

buying certified cars. Buyers who obtained more $SPRICE$ than $CPRICE$ were more likely to

report buying their particular vehicle due to greater satisfaction with their success in

negotiating and financing their vehicle (see Table 3.6b). More buyers who switched to

certified cars upon obtaining $SPRICE$ (vs. $CPRICE$) said that they would not pay more for a

certified car (68% vs. 57%), suggesting that they were less likely to attribute switching to

obtaining higher quality. Among all certified car buyers, however, we observe that those who

perceived that their vehicle was better quality than non-certified cars paid *more* premium

than the average premium that buyers were willing to pay for certified cars ($467 more vs.

$106 more).

Buyers' responses to the reasons for switching to the type of used vehicle that they

purchased, when crossed with the information they obtained, provides additional insight that

is consistent with our theorizing above. Thus, while both types of product (price) information

lead consumers to be less (more) likely to buy certification; our results suggest that the

buyers are influenced in different ways. We are however unable to formally include these

110

factors into our empirical model due to the modest sample size of these subgroups, and leave it to future research to study them.

Thus far, our findings highlight the important role of online information obtained by buyers in influencing key purchase outcomes. In our second post-hoc analyses, we further investigate whether consumers' shopping or purchase goals affected their likelihood of obtaining different types of information. Specifically, we examine if there were any significant differences in outcomes for consumers whose original search set consisted of *only* used vehicles (*Want Used*) compared to consumers who were open to purchasing new and used vehicles (*Open*). We observe that in the absence of the four types of online information, the buyers who *Want Used* have a lower baseline propensity to buy certification ($\mu = 0.32$, s.e.$= 0.01$) than *Open* buyers ($\mu = 0.40$, s.e. $= 0.02$), which is significant at $p<0.001$. Ceteris paribus, we find that buyers who were *Open* also obtained more comparative than specific information. Thus price information *reinforces* their propensity to choose certified cars, while product information has an opposite effect, *decreasing* their propensity to buy a certified used car. In contrast, for buyers who only *Want Used* cars, product information *reinforces* their likelihood of *not* buying certified cars. However, *Want Used* buyers who obtained price information are more likely to purchase certification.

Prior to discussing the implications of our findings it is important to acknowledge the limitations of this work. First, we are limited by our reliance on secondary data collected by a third party. However, this detailed data set collected by one of the largest market research firms in the US represents one of the most extensive surveys of used vehicle buyers and the measures used possess good psychometric properties. Second, common methods bias is

111

mitigated to a large extent by having each response correspond to vehicle registrations and tied objectively to a verified purchase. Yet, this remains a possibility. The 3SLS model exploits information regarding the correlations of the residuals across different equations in the system and therefore is theoretically more asymptotically efficient if there are common unobservables that affect all dependent variables. However, we cannot completely rule out the possibility that some of the observed differences in estimates across 3SLS and OLS could be due to misspecification of the instruments rather than superiority of simultaneous equations as an estimation approach.

3.6 Discussion

3.6.1 Implications

Certified pre-owned programs help manufacturers keep used-car residual values high and create vehicles with higher resale values. Certified used cars are also believed to be more profitable to dealers. Consequently, manufacturers as well as dealers have a strong incentive to promote certified used cars. As for consumers, certification may increase aggregate consumer surplus by increasing the average quality of cars traded in the used vehicles market. Certification also expands the market by making luxury brand vehicles affordable to consumers that would have otherwise not been able to purchase them. However, since such certification is done by the manufacturers/dealers themselves, the value of such certification to consumers has been questioned. The presence of alternate mechanisms to lower asymmetry adds to this debate about the value of certification in the market for used cars. Our findings show that even after controlling for a wide range of potentially confounding variables, certified cars commanded a premium, suggesting that consumers have a positive valuation for certification. This premium may be explained by several factors. The results in

112

Table 3.3 highlight significant differences between the population of certified vehicles and non-certified ones. For instance, certified vehicles were more likely to be low mileage with lower variance in usage (t = 6.97, *p < .01*), newer model year (t = 4.20, *p < .01*), and luxury makes (t = 8.52, *p < .01*). This suggests that consumers might benefit from the selective culling of certified used cars, which might be particularly valuable to risk-averse consumers as it allows them to enter the market for used vehicles. Thus buyers who might otherwise not consider purchasing a used vehicle might be able to purchase a certified used vehicle (Vella 2006). We also find that consumers who were more satisfied (reported as >8 on a scale of 1-10 vs. those less satisfied) with their certified vehicle were significantly more likely to: recommend their make-model to others (79% vs. 54%, p < .001), purchase a new vehicle of same make in the future (42% vs. 27%, p < .001), and purchase certified vehicles again (58% vs. 34%, p<0.001). Satisfied consumers who bought certified cars were also more likely to return to their dealer for post-purchase services than were similarly satisfied counterparts who purchased non-certified cars (83% vs. 66%), highlighting the additional benefits dealers get from certification.

While certification, as seen above, plays an important role in the used vehicle market, one of the most significant developments in auto-retailing has been the dramatic increase in the amount and variety of online information available to consumers. However, the impact of such decentralized online information on consumer purchase behavior and choices has not been examined before. We find that, after controlling for detailed vehicle, buyer and market characteristics, buyer pre-purchase vehicle consideration sets, as well as offline information search, buyers' *value for certification* is significantly impacted by information retrieved from

online sources. However, interestingly this impact depends on the type of information

obtained (see Figure 3.1).

	Comparative Product	Comparative Price	Specific Product	Specific Price
Certification Choice	–	+	–	+
Non-Certified Price	+	–	+	–

Figure 3.1: Impact of Online Information on Choice of Certification and Price of Used

Cars

While vehicle specific product information can help to reduce the knowledge

uncertainty for consumers seeking used cars, comparative product information about the

product class, including data on vehicle speed, handling, and road performance, and

reliability, can help reduce performance uncertainty. We find that these two categories of

online information – *specific and comparative product* information – reduce the likelihood of

a buyer purchasing a certified used car, while significantly increasing the price paid for non-

certified used cars. Thus, from a seller's perspective, (comparative and specific) product

information is a substitute to traditional certification, and a complement to the sale of non-

certified used cars. On the other hand, vehicle/transaction specific price information can

reduce the financial uncertainty, while comparative information about the price of new

alternatives can help reduce economic uncertainty about the value of the purchase. We find

that *specific* as well as *comparative price* online information increases buyers' demand for

certification, while reducing the price paid for non-certified used cars. Thus, from a seller's

perspective, the two categories of online price information *complement* traditional

certification, and act as a substitute for the sale of non-certified used cars. These findings

have significant implications for traditional dealers of used cars. Sellers of certified used cars

would benefit from providing consumers with easier and inexpensive access to

114

transaction/price related information, while sellers of non-certified used cars would benefit from providing consumers with greater access to information about specific product features and product information about new alternatives.

Our findings also highlight implications for the strategic partnerships between traditional dealers and online information providers. Since online information providers vary in the type of information they provide, partnering or linking with the right online information providers would be mutually beneficial to dealers as well as online infomediaries. For instance, dealers of certified used cars would benefit from affiliations with online infomediaries such as Capital One Auto Finance, and E-loans.com that provide transaction information. In addition, they would also benefit from partnering with online sites that provide information on *both* used and new vehicles. As indicated by our results, buyers who obtained comparative price information were more likely to purchase certification, highlighting the interrelationships between new and certified-used car sales. Alternatively, dealers selling non-certified used cars (for instance, dealers of non-luxury and high-mileage used cars), as well as consumers, would benefit from dealers' affiliation with online infomediaries such as AutoSafety.org, LemonaidCars.com, and CarFax.com that provide comparative product information that serve as a substitute to certification but complement non-certified used car purchases. Identifying the right online information partners would not only help traditional dealers target the right customer segments but also optimize their inventory of certified and non-certified used cars. Concomitantly, online infomediaries would also benefit by better highlighting the value of their information in reducing asymmetries in such markets.

We also find in our post-hoc analyses that consumers who begin their automobile search with different purchase intentions differ in the impact online information has on their choice of certification. The online medium makes it possible for sellers to understand consumers' underlying search deliberations. For example, knowing that consumers are open to both new and used vehicles, the used car seller can increase consumers' likelihood of buying certification by offering comparative price information about new car alternative to highlight the value of purchasing certification. On the other hand for consumers whose consideration sets are limited to used cars, a seller can increase her revenue on non-certified used cars sales by providing the right product information.

Our findings relating to buyer characteristics and their likelihood of obtaining online information have interesting implications. A recent survey conducted by Automotive Retailing Today, a coalition of automakers and dealers whose goal is to narrow the gap between media accounts of dealership conditions and consumers' experiences, finds that the majority of the minority buyers that were surveyed said that their dealership did not give them enough information to make an informed purchase, and that the dealerships often did not honor their commitments (Harris 2005). Our results indicate that minorities as well as less-educated consumers are also less likely to obtain the various categories of online information prior to their purchase. This has important implications for their welfare, as these are typically the same consumers who tend to be discriminated against by traditional dealers. Online information intermediaries can add greater value to these consumers who are more prone to discrimination in traditional channels. Currently, "the 'Car Buyer's Bill of Rights' mandates dealers to reveal vehicle history along with a copy of the inspection report when selling certified used vehicles and provide a two-day sales contract cancellation policy"

(CIRP 2007). However, our findings suggest the need for stronger public policy measures to ensure greater transparency in transactions with disadvantaged consumers.

Our results relating to the impact of online information on price of used cars also extend earlier findings. For instance, Zettelmeyer et al. (2006), find that online buyers paid on an average about 2% less than offline buyers for vehicles. In our study, by teasing out different types of online information sought by used vehicle buyers, we obtain more nuanced effects of online channel use. Another interesting finding relates to consumers' use of online and offline sources of information. We find that while impersonal/ commercial sources of offline information (e.g., classifieds in TV/magazines/radio) complement online information search, the use of personal information sources (e.g., friends and relatives) serves as a substitute to online information search in the context for used cars – a likely indication of the importance of trust in the purchase of used goods. This suggests that dealers of used goods might benefit from cost-effective alternate quality signals such as reputation mechanisms and ratings from earlier transactions to engender greater trust in consumers.

3.6.2 Conclusion

Secondary or used good markets are an important part of the economy and have been growing rapidly in many product categories. Clearly, secondary markets are an important category for vehicle manufacturers and play an important role in the demand as well as the profitability of new cars for manufacturers, as well as dealers. The rapid growth of the Internet and decentralization of information has dramatically changed the balance of power between consumers and car dealers. Given that used cars are twice as profitable for dealers as new vehicles (CIRP 2007), understanding the impact of the changes brought about by transformations in the informational landscape becomes paramount. While earlier studies

have examined the impact of the Internet on the market for new cars, there have been very few studies of secondary markets in general, and more specifically the impact of the Internet on the market for used cars. Our study is among the first to examine the cross-channel impacts of different types of online information on the traditional market for used cars. In addition, we focus on the impact of the decentralization of online information on consumers' value for a relatively centralized signal -certification on used cars. Our study highlights interesting relationships between different types of online information and consumers' value for certification, and points to the need to disentangle these effects empirically to better understand their differential impacts on the outcomes of interest to buyers and sellers. More broadly, our findings about the impact of the different types of online information on consumer demand for traditional quality signals and price outcomes provide useful guidelines for other secondary markets with disclosure, and also perhaps, markets for certified products that rely on quality labeling.

While the primary focus of this study has been on the impact of online information on certification, other mechanisms such as seller quality, warranties, product guarantees, and branding, also serve to reduce information asymmetries in many markets. It would be useful to examine the impact of the increased availability of online information on alternate quality signaling mechanisms. In addition, while this study examines the impact of online information on consumer choices in a classic offline lemons market, it would be interesting to study the effect of such decentralized information on quality signaling mechanisms in online secondary markets such as EBay and Amazon Marketplace, which have gained prominence. This study is a first step in understanding how the Web and digitization have transformed the informational landscape for consumers from one reliant on centralized

sources of information to one supported by distributed/ decentralized information. The implications of this development for consumers and marketers across other markets remains to be better understood.

Table 3.1. Operationalization of Constructs

Construct	Operationalization
Vehicle Price	Total Price in $ (excl. tax, license, trade-in)
Certification	1 – Purchased certified vehicle; 0 – otherwise
Model Year	Vehicle Model Year (1999 up to 2004) with base year 1998
Mileage	Miles at Purchase /1000
Luxury Make	1 – If Luxury nameplate; 0 – otherwise
Annual Income	15 ordered intervals ranging from less than $25K to over $250K (pre-tax income)
Minority Race	1 – If African American or Hispanic; 0 – otherwise
Low Education	1 – If less than a high school graduate, 0 – If more educated
Gender	1 – Male; 0 – Female
Age	Age in Years
Married	1- Yes; 0- No
Previous Used Car	1 – Has previously owned a used car; 0- otherwise
Pre-purchase intention	Dummies for Want Used, Open, and Want New
Use offline classifieds/ads	Average of 2 items: use of offline classifieds/ads to locate and research vehicle 0-1
Use offline personal sources	Average of 2 items: use of prior experience and personal recommendations to locate and research vehicle 0-1
Visit Online Before Dealer	1 - conducted online search prior to visiting dealers; 0- otherwise
Comparative Product CPROD	Average of 4 items corresponding to online information found on a scale of 0-1
Comparative Price CPRICE	Average of 3 items corresponding to online information found on a scale of 0-1
Specific Product SPROD	Average of 3 items corresponding to online information found on a scale of 0-1
Specific Price SPRICE	Average of 4 items corresponding to online information found on a scale of 0-1
Seller Type	New car dealer or Used car dealer
Market size	Dummies for Rural, small, metro
Vehicle History Reports	1- If the buyer had access to a vehicle history report prior to purchase, 0 otherwise
Remaining OEM warranty	1- if vehicle had remaining original/manufacturer warranty; 0- otherwise
Additional Warranty	1 – Purchase additional warranty; 0 – otherwise
Satisfaction with vehicle overall quality	10 point scale for overall rating of vehicle
Post-Purchase Defects	Number of problems encountered with vehicle after purchase

Table 3.2. Factors for Online Information Search

Please tell us whether you found this information while searching on the Internet (Yes/ No)

Information Found While Searching Online	F1:CPROD	F2:CPRICE	F3:SPROD	F4:SPRICE
Road tests/articles about vehicles	**0.81**	0.19	0.16	0.11
Performance data on vehicles (speed, handling, etc.)	**0.79**	0.10	0.26	0.10
Vehicle reliability information	**0.75**	0.21	0.10	0.11
Safety information	**0.70**	0.08	0.23	0.22
Price of used vehicles	0.13	**0.89**	0.11	0.05
Trade-in values	0.12	**0.87**	0.10	0.03
Prices of new vehicles	0.14	**0.81**	0.12	0.24
Dealer cost/invoice of new vehicles	0.10	**0.62**	0.09	0.28
Used vehicle photographs	0.20	0.11	**0.89**	0.25
Size and dimensions of vehicle	0.17	0.12	**0.88**	0.13
Options and features available on used vehicles	0.20	0.14	**0.82**	0.15
Locate used vehicles for sale	-0.06	0.00	**0.65**	0.19
Tool for calculating monthly payments	0.11	0.07	0.21	**0.88**
Service contract/extended warranty information	0.14	0.18	0.22	**0.87**
Special financing/discount offers	0.05	0.15	0.08	**0.78**
Information on certified used vehicles	0.15	0.01	0.19	**0.75**
Factor Reliability (Cronbach alpha)	**0.84**	**0.86**	**0.85**	**0.86**

Table 3.3. Differences Across Non-certified and Certified Samples

Variables	Non-Certified Cars (N =2094)		Certified Cars (N = 1119)		t- Statistics (df = 3211)
Vehicle Characteristics					
Price paid ($1000)	14.83	(7.26)	*17.62*	*(7.88)*	**t = -10.06**
Miles on Vehicle (000's)	*35.82*	*(20.75)*	30.78	(17.00)	**t = 6.97**
Model year	2.87	(1.35)	*3.07*	*(1.31)*	**t = -4.20**
Luxury	0.20	(.40)	*.34*	*(0.47)*	**t = -8.52**
Consumer Demographics					
Age	47.58	(14.99)	*49.65*	*(14.76)*	**t = -3.75**
Gender	*.59*	*(.49)*	.56	(.50)	**t = 1.88**
Low Education	.26	(.44)	.23	(.42)	t = 1.80
Income	5.88	(3.32)	*6.40*	*(3.50)*	**t = -4.14**
Minority Race	.06	(.24)	*.10*	*(.29)*	**t = -3.25**
Married	.70	(.46)	*.71*	*(.45)*	t = -0.94
Consumer Experience/ Psychographics					
Previous Used Car	*.63*	*(.48)*	.58	(.49)	**t = 2.79**
Want Used vehicle	*.65*	*(.48)*	.58	(.49)	**t = 4.35**
Open to New and Used	.30	(.46)	*.36*	*(.48)*	**t = -3.46**
Online Information					
CPROD	.33	(.40)	*.38*	*(.41)*	**t = -3.51**
CPRICE	.34	(.39)	*.40*	*(.40)*	**t = -3.87**
SPROD	.44	(.46)	*.50*	*(.44)*	**t = - 3.06**
SPRICE	.16	(.27)	*.24*	*(.32)*	**t = -6.67**
Controls					
Visit online before dealer	.43	(.50)	*.46*	*(.50)*	**t = -1.85**
Offline Classifieds/ads	.56	(1.16)	*.57*	(1.16)	t = -.22
Offline Personal	*.25*	(1.27)	.23	(1.26)	t = .35
Bought from metro market	.16	(.36)	*.19*	*(.39)*	**t = -2.47**
Bought from small market	.42	(.49)	*.48*	*(.50)*	**t = -2.87**
Bought from new vehicle seller	.70	(.46)	*.84*	*(.37)*	**t = -8.72**
Additional warranty	*.82*	*(.38)*	.32	(.47)	**t = 32.90**
Original Manufacturer warranty	.49	(.50)	*.53*	*(.50)*	**t = -2.22**
Obtained history reports	.29	(.46)	*.40*	*(.49)*	**t = -5.96**
Short term defects (problems)	.80	(1.31)	*.81*	(3.22)	t = -.40
Satisfaction with vehicle quality	7.95	(1.92)	*8.24*	*(1.92)*	**t = -4.21**

* $p < .10$, ** $p < .05$, *** $p < .01$; unpaired sample t-tests. Standard deviation is shown in parentheses.

Notes: The greater mean is *italicized*, while significant t-statistics are in **bold**. Our sample is 58% male, 7.50% minority (African Americans/Hispanics), and 25% low education (high school and below) buyers with a median age of 48 years. 75% of the purchases were made with new car dealers. Buyers provided detailed accounts of their valuation and use of vehicle certification programs, and online and offline information search processes. 35% purchased a certified used vehicle, and 61% had previously owned a used car. 56.2% buyers used the Internet to shop for their used vehicle, and spent an average of 7 hours searching 1.96 (SD = 2.53) third-party, 1.42 (SD = 2.58) manufacturer, 1.95 (SD = 3.89) dealerships, .31 (SD = 1.22) newspapers and .05 (SD = .50) chat room/bulletin board websites. 78% of Internet users conducted online research 6 weeks prior to visiting physical dealer locations.

Table 3.4. Value of Certification

	M1a. LnPRICE	M1b. CERTI	M1c. CPROD	M1d. CPRICE	M1e. SPROD	M1f. SPRICE
A. VEHICLE CHARACTERISTICS						
Miles	**-.005**	**.039**	-.000	-.000	-.000	-.000
	(.000)*	**(.005)***	(.000)	(.000)	(.000)	(.000)
Model Year	**.099**	**-.794**	-.008	-.005	-.008	-.001
	(.005)*	**(.096)***	(.006)	(.006)	(.006)	(.005)
Ln Price		**7.939**				
		(.857)*				
B. CERTIFICATION AND ONLINE INFORMATION						
Certification	**.123**					
	(.019)*					
Certification * Miles	.000					
	(.004)					
Certification * Luxury	.018					
	(.066)					
CPROD	**.090**	**-.987**				
	(.023)*	**(.119)***				
CPRICE	**-.084**	**1.314**				
	(.027)*	**(.163)***				
SPROD	**.029**	**-.562**				
	(.016)*	**(.082)***				
SPRICE	**-.093**	**.951**				
	(.020)*	**(.091)***				
CPROD * Certification	-.003					
	(.034)					
CPRICE * Certification	-.000					
	(.033)					
SPROD * Certification	.002					
	(.040)					
SPRICE * Certification	.001					
	(.032)					
C. BUYER DEMOGRAPHICS AND PSYCHOGRAPHICS						
Income	.001	-.024	.001	.003	.000	.001
	(.002)	(.017)	(.002)	(.002)	(.002)	(.002)
Minority	.003	.114	**-.050**	**-.060**	**-.059**	**-.037**
	(.017)	(.169)	**(.020)**	**(.019)***	**(.019)***	**(.017)**
Low education	.014	-.152	**-.048**	**-.040**	**-.055**	-.011
	(.010)	(.109)	**(.013)***	**(.012)***	**(.012)***	(.011)
Gender (male)	.015	**-.239**	-.004	.015	.000	.003
	(.010)	**(.097)**	(.011)	(.011)	(.011)	(.010)
Age	**-.001**	**.007**	-.000	**-.001**	**-.002**	-.000
	(.000)*	**(.004)***	(.000)	**(.000)**	**(.000)***	(.000)
Married	-.015	.089	-.001	-.001	-.005	.005
	(.010)	(.109)	(.013)	(.012)	(.012)	(.011)
Previous Car Used	-.012	**.161**	**-.020**	-.016	-.003	-.011
	(.010)	**(.092)**	**(.011)**	(.011)	(.010)	(.009)
Want Used	**-.063**	**.742**	-.018	**-.028**	**.020**	-.014
	(.024)*	**(.186)***	(.011)	**(.011)***	**(.011)**	(.010)

123

D. CONTROLS: (OFFLINE) INFORMATION SEARCH						
Use of classifieds	-.011	.096	.044	.042	.066	.026
	(.004)***	(.042)**	(.005)***	(.005)***	(.004)***	(.004)***
Use of personal sources	-.013	.118	-.024	-.023	-.033	-.020
	(.003)***	(.036)***	(.004)***	(.004)***	(.004)***	(.004)***
Constant	3.345	-26.240	-.137	-.195	-.112	.005
	(.076)***	(2.998)***	(.092)	(.087)**	(.086)	(.078)
Fit statistics	R^2= .60	R^2=.12	R^2= .49	R^2= .51	R^2= .63	R^2= .30

* $p <.10$,   ** $p < .05$, *** $p < .01$. Standard errors shown in parentheses.

Notes: N = 3213 purchases in all models. The online information factors are multiplied by a factor of 10 in M1a and M1b. All models M1a-M1f contain 125 car dummies for vehicle make-model-trim fixed effects. Additional controls (not shown) common to the price model M1a and choice of certification model M1b are market area (rural, small, metro), type of seller (new vehicle or used vehicle dealer), one of the short term quality variables- number of post-purchase defects, order of online search compared to dealer visits, and consumer psychographics- Want Any. The price model M1a contains additional controls- buyer's satisfaction with overall vehicle quality, availability of vehicle history reports, remaining manufacturer warranty and purchase of additional warranty. Interaction components are centered to reduce multicollinearity. All variables in panel B are modeled as endogenous and estimated using the instrumental variables technique with a surfeit of instruments. Details of the IV regression are provided in the Appendix.

Table 3.5. Comparing the Impacts of Online Information on Price Paid

	M1a. Ln Price: Pooled Sample	M1b. Choice of Certification	M2a. Ln Price: Non-certified Sample	M2b. Ln Price: Certified Sample
CPROD	.090	-.987	.110	-.015
	(.023)***	(.119)***	(.032)***	(.016)
CPRICE	-.084	1.314	-.036	.028
	(.027)***	(.163)***	(.021)*	(.031)
SPROD	.029	-.562	.053	-.013
	(.016)*	(.082)***	(.025)**	(.013)
SPRICE	-.093	.951	-.107	-.001
	(.020)***	(.091)***	(.031)***	(.014)
Observations	3213	3213	2094	1119
Fit statistics	R^2= .60	R^2=.12	R^2= .68	R^2= .86
χ^2	$\chi^2(158) =$ 12407.66***	$\chi^2(148) =$ 347.84***	$\chi^2(151) =$ 6068.98***	$\chi^2(151) =$ 7914.57***

* $p <.10$,   ** $p < 0.05$, *** $p < 0.01$; Standard errors shown in parentheses.

Notes: The online information factors are multiplied by a factor of 10 in M1a and M1b. Model M1a and M1b for the pooled sample are reproduced from Table 4 for comparison purposes. The results from applying 3SLS to the non-certified sub-sample and certified sub-sample are shown in M2a and M2b, respectively.

As noted in the paper, an increase (decrease) in the likelihood of purchasing a *certified used car* corresponds to a decrease (increase) in the likelihood of purchasing a *non-certified used car*. Thus M1b and M2a taken together indicate the complementary/substitutive effects of product and price information on non-certified used cars.

The controls variables included in models M1a-M1b and M2a-M2b are the same as those described in the notes for Table 4. All four online information variables in M2a and M2b are modeled as endogenous and estimated using the instrumental variables technique with a surfeit of instruments. Details are provided in the Appendix.

Table 3.6a. Comparing the Reasons Reported for Purchasing Chosen Vehicle by Consumers Who Obtained Product Information (SPROD vs. CPROD)

Reason	SPROD	CPROD
The exact vehicle I wanted was not Certified	**14%**	1%
Seller had the exact vehicle (color, options, etc) I wanted	**51%**	33%
Obtained more features for the same price	**22%**	8%
Satisfied with the vehicle condition- explanation of features, cleanliness etc.	**7.09**	6.80
Obtained vehicle history reports	40%	**66%**
Rejected vehicles based upon the contents of this report	14%	**25%**

Table 3.6b. Comparing the Reasons Reported for Purchasing Chosen Vehicle by Consumers Who Obtained Price Information (SPRICE vs. CPRICE)

Reason	SPRICE	CPRICE
Seller offered attractive financing	**18%**	13%
Hassle free negotiation	**50%**	44%
Satisfaction with paperwork/finance process	**8.50**	7.90
Special financing offer/discount included with certification	**47%**	28%
Saw/realized the value in Certification	28%	**46%**

Chapter 4: Epilogue

The Internet as an expanding channel for commerce fundamentally changes the way in which parties involved in electronic transactions interact with one another. While this new medium poses some challenges, especially those that stem from information asymmetry between parties, it also allows for the creation and design of new ways to deal with such frictions. One such mechanism, and a key variable in electronic markets, is the availability of online information. This has led to buyers and sellers devising new strategies to manage the acquisition and provision of information, respectively.

The Internet facilitates two-way information exchange, where consumers seek and search for information about products and sellers, and in doing so, are also indirectly offering information about themselves to the firm in the form of their information-seeking behaviors. With the development of new technologies to track and capture consumers' search behaviors, firms today have access to a powerhouse of information about consumers' underlying needs and preferences. My dissertation highlights the importance of studying consumers' information seeking behaviors in the online medium. In the first essay, the information obtained by consumers is centrally and directly provided by the seller, whereas in the second essay, the impacts of decentralized information gathered from several online sources that are available to consumers seeking to purchase durable goods is examined. The findings from these studies show that consumers' search patterns reflect 'meaningful' information about consumers' information needs and preferences. Further, this information is important for firms because it allows them to strategically and optimally interact with customers by customizing the provision of information to suit their needs.

The findings from this dissertation add to the streams of literature that study the impacts of information obtained by consumers in online channels. Additionally, the impacts of product- and price-related information on consumers' choices and outcomes in both the online and offline

markets are separately quantified. Essay 1 shows that online information can impact consumers' decision-making at a very fine-grained level and that it is fruitful to examine the different impacts of information across various decision stages faced by online consumers. I find interesting tradeoffs between the effects of information on purchase outcomes within a session and across sessions for a given consumer. Uncovering consumer segments that respond differently to information interventions is likely to be a useful finding for online firms. e-tailers can leverage this knowledge to fine-tune or micro-customize their online information provision strategies to consumers, and obtain better market outcomes. Essay 2 sheds light on the cross-channel impacts of online information on how consumers value traditional quality signals. As internet markets continue to mature, online consumers today find themselves faced with several quality signals. It is therefore important and useful to study how various sources of information interact with one another. For instance, how does the presence of one signal or source of information impact consumers' valuation of another? This essay takes a step towards answering this important question. It examines the dynamics between a traditionally used signal of quality in secondary markets and information related to the purchase of used goods available today in online channels. The findings reveal interesting patterns of substitution and complementarity among four types of online information and certification. Once again, these results suggest that firms should pay attention to what kinds of information they provide to their consumers; because information leads consumers to make different choices and face different outcomes.

In conclusion, the essays from this dissertation seek to inform firms in designing better ways to interact with consumers' through the provision of online information. As the digitization of markets continues, it also opens new grounds and presents several novel opportunities to sellers- one of them being the management of online information. Sellers that learn to leverage the power of information to help better interact with their consumers in digital markets will be better positioned to succeed, and this dissertation was a step toward helping firms to better understand the impacts of product and price related information.
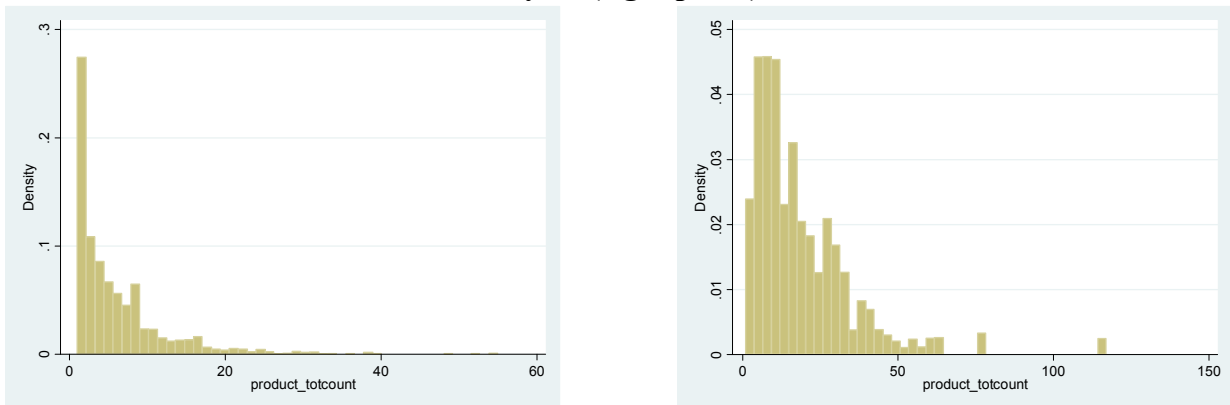
Appendices

APPENDIX A: Observed Behaviors Across Buyers and Non-Buyers
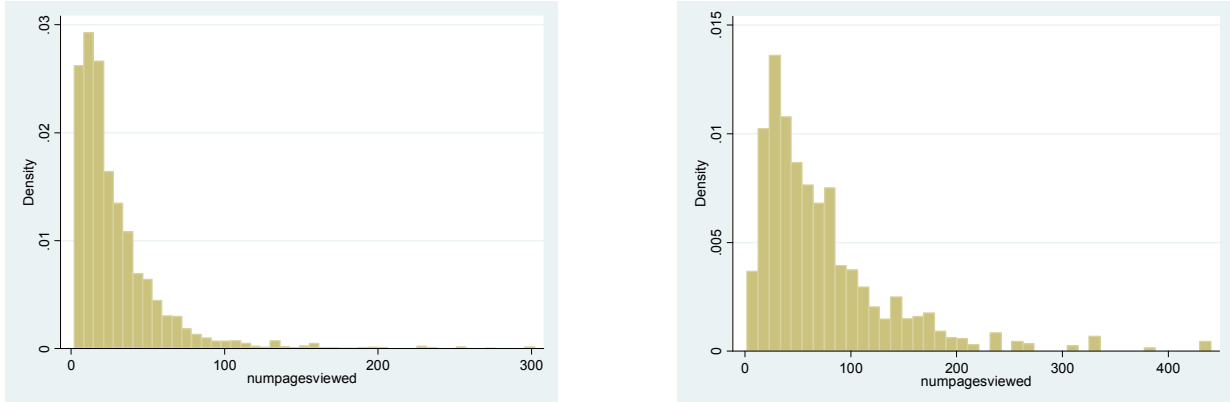
Figure A1a. Distribution of the number of pages viewed in a session across non-buyers (left panel) and buyers (right panel)
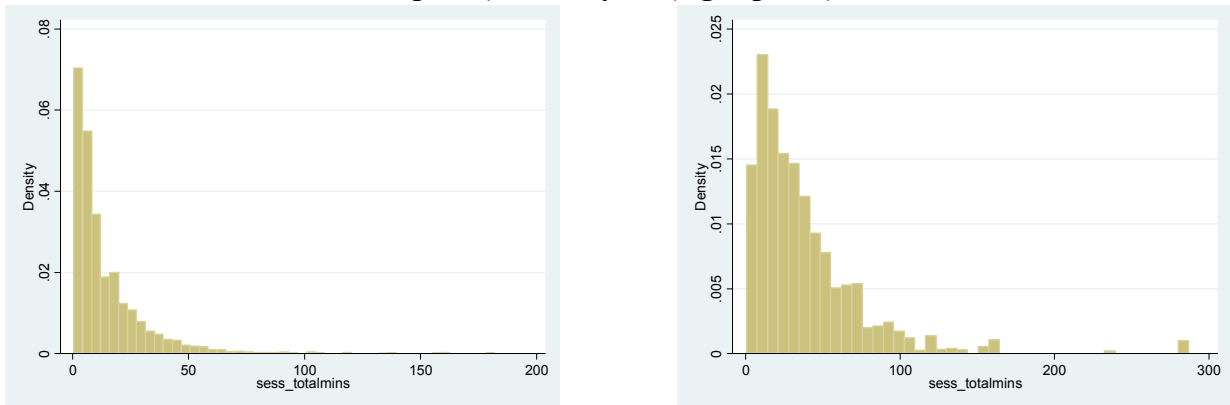
Figure A1b. Distribution of the session length (minutes) across non-buyers (left panel) and buyers (right panel)

Figure A1c. Distribution of the number of products viewed in a session across non-buyers (left panel) and buyers (right panel)

APPENDIX B: Details of the Instrumental Variable Regression

Endogeneity

In the presence of endogeneity, OLS estimates are biased and inconsistent (Maddala1992), whereas properly specified 2SLS/3SLS models are consistent. Further, in the presence of contemporaneous error-covariance, a method like 3SLS that makes use of the cross-equation correlations of the disturbances is asymptotically more efficient than 2SLS, and is the one we use.

Our system of equations (1)-(3) is described in the main paper The 3SLS procedure is as follows: (1) Each endogenous variable in the system (here, Certification, Certification X Miles, Certification X luxury, CPROD, CPRICE, SPROD and SPRICE, and Certification X Online Information, and PRICE) is regressed on all of the exogenous variables in the system, and predicted values of the endogenous variables are calculated in the first-stage. (2) These predicted values are used as instrumental variables for the endogenous variables in the second-stage OLS regressions, from which estimates of error terms and variance-covariance matrices are obtained. (3) Using the cross-equation disturbance correlation estimates, generalized least squares estimation is applied to a single equation representing all the system equations.

Instruments

The price model (1) contains 11 potentially endogenous regressors. We identify a total of 12 instruments: a binary variable-*Shop Certified* -that describes that the buyer intentionally shopped only for a certified vehicle; and interactions of *Shop Certified* with *Miles* and *Luxury*. Further, we use as an instrument buyers' belief that there is *Value* in certification programs (1-10). A set of four variables that measure the buyers' ratings of the importance of each type of information on a scale of 1 to 10, i.e. importance of comparative and vehicle-specific price and product information are used to instrument online information. Finally, the interactions between *Shop Certified* and the four *importance* of online information variables are used as instruments for the interaction terms The certification equation (2) contains 5 potentially endogenous regressors- four online information factors and price. Two variables are used to instrument for price- whether seller offered financing and whether buyer considered the vehicle purchased to be the

best deal for the money. Finally, we use a dummy to indicate whether the buyer shopped for a specific vehicle.

We need at least as many instruments as there are endogenous regressors in each equation. However, while the asymptotic efficiency of the estimation improves as the number of instruments increases, but so does the finite-sample bias (Hahn and Hausman 2002, Johnston and DiNardo 1997). Thus, we chose a parsimonious non-redundant subset of instruments needed to appropriately identify our system, by examining subsets of the orthogonality conditions using difference-in-Sargan or C tests (Hayashi 2000).The instrumental variables regression is estimated using STATA 9.2's ivreg2 and reg3 procedures (see Baum et al. 2003, 2007), and the statistics are reported in Tables A1 and A2.

We first test for the presence of endogeneity in each single equation model (1) and (2) using a form of the Durbin-Wu-Hausman test that is robust to various violations of conditional homoskedasticity (Baum et al. 2007, p.16). The test statistic is rejected for both Price [1] ($X^2(11) = 26$ (p<0.00)) and for Certification (2) ($X^2(5) = 36.19$ (p<0.00)), suggesting the need to account for endogenous regressors.

We also find evidence of the presence of significant heteroskedasticity in our models using the Pagan and Hall test statistic (appropriate for IV estimation given sufficient sample size): for Price, $\chi^2(1) = 11.73$ (p<0.00) and for Certification, $\chi^2(1) = 58.17$ (p<0.00). We therefore estimate equations (1) and (2) using heteroskedasticity-robust IV techniques with sandwich V-C matrices. Next, we provide details of the diagnostics/tests conducted to ensure the validity and relevancy of our instruments.

Validity and relevance of instruments

Good instruments must be both valid and relevant to ensure that the model is identified. We examine the validity of the overall set of instruments by examining the orthogonality (exogeneity) of the instruments to the structural equation (Price or Certification) using a heteroskedasticity-robust overidentification test (numerically equivalent to Hansen J as shown by Baum et al. 2003). This is an omnibus test, with the null hypothesis that all the excluded exogenous variables, the instruments, are uncorrelated to the regression error in the main equation. This statistic is chi-square distributed with degrees of freedom equal to the difference between the number of instruments and the number of endogenous variables. If rejected, IV

estimates will be inconsistent. We however fail to reject this statistic for both Price ($\chi^2(1) = 1.42$ (p=0.24)) and Certification ($\chi^2(2) = 2.92$ (p=0.23)) equations, ensuring that the instruments are valid or orthogonal to the residuals in the structural equations.

Additionally, we want our instruments to be relevant, or significantly correlated with the endogenous regressors they replace. This is a test of the rank condition that each of K canonical correlations (K= endogenous regressors) is different from zero. In Tables A1 and A2, we present the partial R^2, Shea partial R^2 (which takes into account correlations among instruments), and the first-stage F-statistics. For a model with a single endogenous regressor, an F below 10 is a cause for concern as shown by Staiger and Stock (1997). However, in models with multiple endogenous regressors, as is the case in equations [1] and [2], additional diagnostics are required (Stock and Yogo 2005). We therefore also examine the Kleibergen-Paap statistics for under-identification, which is appropriate under the presence of heteroskedasticity (Kleibergen and Paap 2006). The Kleibergen-Paap rk LM statistic (robust versions of the Anderson LM based on null that the smallest canonical correlation is not different from zero) tests the null that the model is of rank K-1, i.e. underidentified. We reject the null in both equations for Price (LM $\chi^2(2) = 110.24$, p<0.00) and Certification (LM $\chi^2(3) = 116.00$, p<0.00), suggesting that the models are not under-identified.

As discussed in several papers in recent econometrics literature, under-identification is not a sufficient shield from weak-identification which causes additional problems of inference (Hahn and Hausman 2002, Staiger and Stock 1997, Stock et al. 2002). Weak identification occurs when the correlation between the instruments and the endogenous variables is small. In such a case, it magnifies the effect of any correlation between the instruments and the error in the original structural equation, and leads to inconsistent IV estimates (Bound et al. 1995). Additionally, as the first stage R^2 approaches zero, finite sample results may differ substantially from asymptotic theory (causing greater bias in IV). Thus Stock and Yogo (2005) provide two tests to assess the effects of weak instruments based on maximal relative bias of IV and test size. The first is a test of the bias in IV compared to the bias in OLS. The second test is concerned with the performance of the Wald test (of the null that β, the coefficient of the

endogenous regressor in the structural equation is zero), which rejects too often when identification is weak. Rejection of the null provides an estimate of the IV rejection rate for β when the true rejection rate is 5% (see Baum et al. 2007, p.24). The corresponding Kleibergen-Paap rk F-statistic for equation [1] is 12.79 and for equation [2] is17.81.

Finally, we assess the joint significance of multiple endogenous variables in our structural equations [1] and [2] using the Anderson-Rubin test (1949). This test is robust to weak instruments (i.e. is less likely to reject in their presence). The null hypothesis that the endogenous regressors are not significant is rejected in equation [1] ($\chi^2(12) = 61.61$, $p<0.00$) and [2] ($\chi^2(7) = 49.95$, $p<0.00$). These results indicate that the endogenous regressors are significant in their respective structural equations. The final IV estimates are provided in Table 5 in the main paper.

Table B1. Instrument validity and relevance statistics for price equation (1) Heteroskedadticity-Robust Tests

Model	Certi	Certi Miles	Certi_ Lux	CPROD	CPRICE	SPROD	SPRICE	Certi X CPROD	Certi X CPRICE	Certi X SPROD	Certi X SPRICE
Partial R^2	.28	.55	.47	.61	.55	.41	.51	.16	.14	.14	.19
Shea partial R^2	.26	.52	.36	.56	.52	.41	.35	.12	.12	.10	.10
First stage F (12, 3053)	95.99 ***	434.91 ***	252.68 ***	259.56 ***	247.82 ***	118.17 ***	242.41 ***	37.17 ***	31.68 ***	37.60 ***	28.70 ***

*** $p<0.001$ ** $p<0.05$ * $p<0.10$

Notes: Structural model: N=3213, Excluded Instruments = 12, Endogenous Regressors = 11, F(158, 3054) = 143.41***, $R^2 = 0.84$

Test for heteroskedasticity in fitted values- Pagan Hall test statistic: $\chi^2(1) = 11.67$***

Under-identification test: Kleibergen–Paap rk LM statistic $\chi^2(2) = 111.04$***

Weak identification test Kleibergen–Paap rk F-statistic = 12.79

Over-identification test Hansen J: $\chi^2(1) = 1.40$ (p=0.24)

Joint significance of endogenous regressors in structural equation Anderson–Rubin Wald test $\chi^2(12) = 61.61$***

Endogeneity of endogenous regressors: DWH test $\chi^2(11) = 26.37$***

Table B2. Instrument validity and relevance statistics for certification equation (2)

Model	Ln Price	CPROD	CPRICE	SPROD	SPRICE
Partial R^2	0.05	0.40	0.40	0.50	0.24
Shea partial R^2	0.04	0.28	0.25	0.21	0.10
First stage F (7, 3062)	20.27 ***	169.98 ***	180.14 ***	263.54 ***	85.98 ***

*** $p<0.001$ ** $p<0.05$ * $p<0.10$

Notes: Structural model: N=3213, Excluded Instruments = 7, Endogenous Regressors = 5, F(148, 3064) = 4.69***, $R^2 = 0.84$

Test for heteroskedasticity in fitted values- Pagan Hall test statistic: $\chi^2(1) = 58.27$***

Under-identification test: Kleibergen–Paap rk LM statistic $\chi^2(3) = 116.01$***

Weak identification test Kleibergen–Paap rk F-statistic = 17.81

Over-identification test Hansen J: $\chi^2(2) = 2.92$ (p=0.23)

Joint significance of endogenous regressors in structural equation Anderson–Rubin Wald test $\chi^2(7) = 49.95$***

Endogeneity of endogenous regressors: DWH test $\chi^2(5) = 36.31$***

Bibliography

Abadie, A. and Imbens, G. 2002. Simple and Bias-Corrected Matching Estimators for Average Treatment Effects, *NBER Technical Working Papers*.

Adomavicius, G., Sankaranarayanan, R., Sen, S., and Tuzhilin, A. 2005. Incorporating contextual information in recommender systems using a multidimensional approach, *ACM Transactions on Information Systems* 23, 103–145

Agarwal, R., and Venkatesh, V. 2002. Assessing a firm's Web presence: A Heuristic evaluation procedure for the measurement of usability, *Information Systems Research* 13(2), 2002, 168–186

Akerlof, G. A. 1970. The Market for Lemons: Quality Uncertainty and the Market Mechanism. *Quarterly Journal of Economics,* **84**(3), 488-500.

Alba, J., Lynch, J.G., Weitz, B., Janiszewski, C., Lutz, R., Sawyer, A., and Wood, S. 1997. Interactive Home Shopping: Consumer, Retailer, and Manufacturer Incentives to Participate in Electronic Marketplaces. *Journal of Marketing,* **61,** 38-53.

Albano, G.L. and Lizzeri, A. 2001. Strategic Certification and Provision of Quality, *International Economic Review,* 42(1), 267-283

Aldrich, J.H., and Nelson, F.D. 1984. *Linear Probability, Logit and Probit Models.* Sage University Press Paper Series, Quantitative Applications in the Social Sciences. Beverly Hills, CA.

Allenby, G.M. and Rossi, P.E. 1999. Marketing Models of Consumer Heterogeneity, *Journal of Econometrics*, 89 (March/April), 57–78.

Anderson, T. W., and Rubin, H. 1949. Estimation of parameters of a single equation in a complete system of stochastic equations, *Annals of Mathematical Statistics,* **20,** 46-63

Andrews, R. L., and Currim, I.S. 2003. A Comparison of Segment Retention Criteria for Finite Mixture Logit Models, *Journal of Marketing Research*, 40, 235-243

Andrews, R. L., Ansari, A. and Currim, I.S. 2002. Hierarchical Bayes Versus Finite Mixture Conjoint Analysis Models: A Comparison of Fit, Prediction, and Partworth Recovery, *Journal of Marketing Research*, 39 (February), 2002, 87-98.

Angrist, J. D., and Guryan, J. 2008. Does teacher testing raise teacher quality? Evidence from state certification requirements. *Economics Education Review*, 27(5), 483-503

Angrist, J. D., and Krueger, A. B. 2001. Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. *Journal of Economic Perspectives,* **15**(4), 69-85.

Ansari, A., and Mela, C. 2003. e-Customization, 40 (May), 131-145.

Ariely, D., Loewenstein, G., and Prelec, D. 2003. Coherent arbitrariness: Stable demand curves without stable preferences, *Quarterly Journal of Economics,* 118, 73-105.

Baum, C. F., Schaffer, M. E., and Stillman, S. 2003. Instrumental variables and GMM: Estimation and Testing. *Stata Journal,* **3,** 1–31.

Baum, C. F., Schaffer, M. E., and Stillman, S. 2007. Enhanced Routines for Instrumental variables/GMM Estimation and Testing. *Boston College Economics Working Paper No. 667.*

Baye, M. R., Morgan, J., and Scholten, P. 2003. The Value of Information in an Online Consumer Electronics Market. *Journal of Public Policy and Marketing,* **22** (1), 17-25.

Blattberg, R.C., and Wisniewski, K.J. 1989. Price-Induced Patterns of Competition, *Marketing Science,* 8(4), 291-309

Bloch, P. H. and Bruce, G.D. 1984. Product Involvement as Leisure Behavior, in *Advances in Consumer Research*, Vol. 11, ed. Thomas C. Kinnear, Ann Arbor, MI: Association for Consumer Research, 197- 202.

Boulding, W., Kalra, A. Staelin, R., and Zeithaml, V.A. 1993. A dynamic process model of service quality: From expectations to behavioral intentions, *Journal of Marketing Research*, 30 (1), 7-27.

Boulding, W., Kalra, A., and Staelin, R. 1999. The Quality Double Whammy, *Marketing Science*, 18(4), 463-484

Boulding, W., Lee, E., and Staelin, R. 1994. Mastering the mix: Do advertising, promotion and sales force activities lead to differentiation?, *Journal of Marketing Research,* 31, 159–172.

Bozdogan, H. 1987. Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions, *Psychometrika*, 52(3), 345–370.

Bradlow, E., Huber, J., Johnson, E., Khan, B., Little, J., and Schkade, D. 1999. Agents to the rescue? *Marketing Letters*, 10(3), 285–300.

Brown, J. R., and Goolsbee, A. 2002. Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry, *Journal of Political Economy,* **110**(3), 481-507.

Bucklin, R.E., and Sismeiro, C. 2003. A Model of web Site Browsing Behavior Estimated on Clickstream Data, *Journal of Marketing Research*, 40(August), 249-267

Bucklin, R.E., and Sismeiro, C. 2004. Modeling Purchase Behavior at an E-commerce Web Site: A Task Completion Approach, *Journal of Marketing Research*, (August), 2004, 306-323

Bunkley, N. 2009. Auto Sales Fall Back as Clunker Program Ends. Accessed at http://www.nytimes.com/2009/10/02/business/02auto.html. Oct 1, 2009.

Busse, M., Silva-Risso, J., and Zettelmeyer, F. 2006. $1000 Cash Back: The Pass-Through of Auto Manufacturer Promotions, American Economic Review, 96(4), 1253-1270

Busse, M., Simester, D., and Zettelmeyer, F. 2010. The Best Price You'll Ever Get: The 2005 Employee Discount Pricing Promotions in the U.S. Automobile Industry, *Marketing Science*, 29(2), 268-290

Bound, J., Jaeger, D. A., and Baker, R. 1995. Problems with Instrumental Variables Estimation when the Correlation between the Instruments and the Endogenous Explanatory Variables is Weak. *Journal of the American Statistical Association*, **90**(430), 443–50.

Canter, D., Rod, R.., and Storrs, G. 1985. Characterizing user navigation through complex data structure, *Behaviour and Information Technology*, 4 (2), 93-102

Catledge, L.D and Pitkow, J.E. 1995. Characterizing browsing strategies in the World-Wide Web, *Computer Networks and ISDN Systems,* **27** (6), 1065–1073.

Chatterjee, P., Hoffman, D.L., and Novak, T.P. 2003. Modeling the Clickstream: Implications for Web-Based Advertising Efforts, *Marketing Science*, 22(4), 520-541.

Chernev, A. 2003. When More Is Less and Less Is More: The Role of Ideal Point Availability and Assortment in Consumer Choice, *Journal of Consumer Research*, 30 (September), 170–83.

Childers, T.L., Carr, C.L., Peck, J., and Carson, S. 2001. Hedonic and utilitarian motivations for online retail shopping behavior, *Journal of Retailing*, 77, 511-533

Chintagunta, P. K. 1993. Investigating the sensitivity of equilibrium profits to advertising dynamics and competitive effects, *Management Science* 39(9), 1146-1163.

CIRP. 2007. The Anatomy and Physiology of the Used Car Business". Research report. *Car Internet Research Program II*, University of Ottawa (March 15).

Compete Online Shopper Intelligence Study. 2010. Accessed at http://blog.compete.com/2010/02/22/online-shopper-intelligence-study-released/

Cooley, R., Tan, P-N., and Srivastava, J. 1999. WebSIFT: The web site information filter system*, In Proceedings of KDD' 99 Workshop on Web Usage Analysis and User Profiling*, San Diego, CA, Aug. 1999. ACM (Masand and Spilipoulou eds.)

Cox, D. F. 1967. *Risk taking and information handling in consumer behavior*, Harvard University Press, Cambridge, MA.

Cutler, K-M. 2005. Certified Used Cars Come Under Fire: Wave of Lawsuits Claim Widely Offered Guarantees Fail to Disclose Past Problems. *The Wall Street Journal* (July 12).

Danaher, P. J., Wilson, I. W., and Davis, R.A. 2003. A comparison of online and offline consumer brand loyalty, *Marketing Sci*ence, 22, 461–476

Dayton, C. M., and Macready, G. B. 1988. Concomitant-Variable Latent Class Models, *Journal of the American Statistical Association*, 83, 173-178.

Degeratu, A. M., Rangaswamy, A., and Wu, J. 2000. Consumer choice behavior in online and traditional supermarkets: The effects of brand name, price, and other search attributes. *International Journal of Research in Marketing,* **17**(1), 55-78

Dempster, A., Laird, N., and Rubin, D. 1977. Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society, Series B*, 39(1), 1–38.

Dewally, M., and Ederington, L. H. 2006. Reputation, certification, warranties, and information as remedies for seller-buyer information asymmetries: Lessons from the online comic book market. *Journal of Business,* **79**(21), 693-729.

Dewan, S., and Hsu, V. 2004. Adverse Selection in Electronic Markets: Evidence from Online Stamp Auctions, *Journal of Industrial Economics,* **52**(4), 463-590.

Dhar, R. 1997. Consumer Preference for a No-Choice Option, *Journal of Consumer Research*, 24 (September), 215–31.

Dhar, R. and Simonson, I. 2003. The Effect of Forced Choice on Choice, *Journal of Marketing Research*, 40 (May), 146–60.

Dholakia, U. M., and Bagozzi, R. P. 2001. Consumer behavior in digital environments. In J. Wind & V. Mahajan (Eds.), *Digital marketing:Global strategies from the world's leading experts* (pp. 163-200). New York: Wiley

Dias, J.G., and Vermunt, J.K. 2007. Latent class modeling of website users' search patterns: Implications for online market segmentation, *Journal of Retailing and Consumer Services*, 14, 359-368

Dick, A., Chakravarti, D., and Biehal, G. 1990. Memory-based Infererence During Consumer Choice. Journal of Consumer Research, **17**(1), 82-93.

Diehl, K., Kornish, L.J, and Lynch, J.G. 2003. Smart Agents: When Lower Search Costs for Quality Information Increase Price Sensitivity, *Journal of Consumer Research*, 30(1), 56-71

Gelb, B.D., Andrews, D., and Lam, S.K. 2007. A Strategic Perspective on Sales Promotions, *Sloan Management Review,* 48(4), Online Only Article

Ginter, J. L., Young, M. A., and Dickson, P. R. 1987. A Market Efficiency Study of Used Car Reliability and Prices. *The Journal of Consumer Affairs,* **21**(2) 258-276.

Gomory, S., Hoch, R., Lee, J., Podlaseck, M., and Schonberg, E. 1999. Analysis and visualization of metrics for online merchandizing, In *Proceedings of KDD' 99 Workshop on Web Usage Analysis and User Profiling*, San Diego, CA, Aug. 1999. ACM (Masand and Spilipoulou eds.)

Gourville, J. T. 1998. Pennies-a-day: The Effect of Temporal Reframing on Transaction Evaluation. Journal of Consumer Research, **24,** 396-408.

Greene, W. 2003. *Econometric Analysis*, 5th Edition, Prentice Hall.

Greenleaf, E. A. and Lehmann, D.R. 1995. Reasons for Substantial Delay in Consumer Decision Making, *Journal of Consumer Research*, 22 (September), 186–99.

Grossman, S.J. 1981. The Information Role of Warranties and Private Disclosure About Product Quality. *Journal of Law and Economics*, 24(3), 461-483

Gourville, J.T., and Soman, D. 2005. Overchoice and Assortment Type: When and Why Variety BackFires, *Marketing Science* 24(3), 382-395.

Gupta, S. and Chintagunta, P.K. 1994. On Using Demographic Variables to Determine Segment Membership in Logit Mixture Models**,** *Journal of Marketing Research*, 31(1), 128-36.

Hahn, J., and Hausman, J. 2002. Notes on bias in estimators for simultaneous equation models, *Economics Letters*, **75**(2), 237–41.

Harris, D. 2005. Poll: Buyers are unhappy with dealers. *Automotive News*, 79, (January 31).

Häubl, G., and Trifts, V. 2000. Consumer Decision Making in Online Shopping Environments: The Effects of Interactive Decision Aids, *Marketing Science,* 19(1), 4-21.

Hauser, J.R., Urban, G.L., Liberali, G. and Braun, M. 2009. Website Morphing, *Marketing Science,* 28(2), 202-223

Hayashi, F. 2000. *Econometrics*. 1st ed., Princeton, NJ: Princeton University Press.

Heath, T.B., Ryu, G., Chatterjee, S., McCarthy, M.S., Mothersbaugh, D.L., Milberg, S., and Gaeth, G.J. 2000. Asymmetric Competition in Choice and the Leveraging of Competitive Disadvantages, *Journal of Consumer Research*, 27(3), 291-308

Heath, C., J. B. Soll. 1996. Mental Budgeting and Consumer Choice, *Journal of Consumer Research*, **23**(1), 40-52.

Heckman, J. 1979. Sample Selection Bias as a Specification Error, *Econometrica,* **47**(1), 153-161.

Heilman, C., Nakamoto, K., and Rao, A. 2002. Pleasant Surprises: Consumer Response to Unexpected In-Store Coupons. *Journal of Marketing Research,* **2**(May), 242–252.

Hodkinson, C., and Kiel, G. 2003. Understanding Web Information Search Behavior: An Exploratory Model, *Journal of End User Computing,* 15(4), 27-48.

Hoffman, D. and Novak, T. 1996. Marketing in hypermedia computer-mediated environments: conceptual foundations, *Journal of Marketing*, 60(July), 50–68.

How Much Does Certification Cost? 2008. Accessed online http://www.cars.com/go/advice/shopping/cpo/stories/story.jsp?story=value (May 12).

iCrossing. 2010. *How America Searches: Online Retail Report 2010* http://www.icrossing.com/articles/How%20America%20Searches%20-%20Online%20Retail.pdf

Inman, J.J., Dyer, J.S., and Jia, J. 1997. A Generalized Utility Model of Disappointment and Regret Effects on Post-choice Evaluation, *Marketing Science*, 16(2), 97-111

Ippolito, P. M., and Mathios, A. D. 1990. Information, Advertising and Health Choices. *RAND Journal of Economics,* **21**(3), 459-480.

iProspect Report. 2007. *Social Networking User Behavior Study*, Accessed at www.iprospect.com/about/researchstudy_2007_socialnetworkingbehavior.htm

Iyengar, S.S. and Lepper, M. R. 2000. When Choice Is Demotivating: Can One Desire Too Much of a Good Thing, *Journal of Personality and Social Psychology*, 79 (6), 995–1006.

Jain, D.C., and Vilcassim, N.J. 1991. Investigating Household Purchase Timing Decisions: A Conditional Hazard Function Approach, *Marketing Science*, 10(Winter), 1-23.

Janiszewski, C. 1998. The Influence of Display Characteristics on Visual Exploratory Search Behavior, *Journal of Consumer Research*, 25 (3), 290-301.

J. D. Power and Associates. 2003. *Used Vehicle Sales and Certification Study*. Accessed at http://www.jdpower.com.au/corporate/news/releases/pressrelease.asp?ID=2003036 (June 4)

J. D. Power and Associates. 2006. *Autoshopper.com Study: Online Used-Vehicle Classifieds Outpace Print Ads Two to One*. Accessed at www.jdpower.com/corporate/news/releases/pressrelease.asp?ID=2006076 (May 25)

Janakiraman, N., Meyer, R. J., and Morales, A. C. 2006. Spillover Effects: How Consumers Respond to Unexpected Changes in Price and Quality, *Journal of Consumer Research*, **33,** 361-369

Janakiraman, R., Sismeiro, C., and Dutta, S. 2009. Perception Spillovers Across Competing Brands: A Disaggregate Model of How and When, *Journal of Marketing Research*, 46, 467-481

Jin, G. Z., A. Kato. 2006. Price, Quality and Reputation: Evidence from an Online Field Experiment, *RAND Journal of Econom*ics, **37**(4), 983-1005.

Jin, G. Z., Kato, A., and List, J. A. 2010. That's News to Me! Information Revelation in Professional Certification Markets, *Economic Inquiry,* 48(1), 104-122.

Jin, X., Zhou, Y., and Mobasher, B. 2004. Web Usage Mining Based on Probabilistic Latent Semantic Analysis, *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (KDD'04), Seattle, August 2004

Johnson, E. J., Moe, W.W., Fader, P. S., Bellman, S., and Lohse, G. L. 2004. On the Depth and Dynamics of Online Search Behavior, Management Science, 50(3), 2004, pp. 299-308.

Johnston, J., and DiNardo, J. 1997. *Econometric Methods*. 4th ed., New York, NY: McGraw-Hill.

Jovanovic, B. 1982. Truthful Disclosure of Information, *Bell Journal of Economics: The RAND Corporation*, 13(1), 36-44

Kamenica, E. 2008. Contextual Inference in Markets: On the Informational Content of Product Lines, *American Economic Review*, 98(5), 2127-2149

Kaul, A. and Wittink, D.R. 1995. Empirical Generalizations About the Impact of Advertising on Price Sensitivity and Price, *Marketing Science*, 14(3), 151-160

Kivetz, R., Netzer, O., and Srinivasan, V. 2004. Alternative Models for Capturing the Compromise Effect, *Journal of MarketingResearch*, 41(August), 237–57.

Kleibergen, F., and Paap, R. 2006. Generalized reduced rank tests using the singular value decomposition, *Journal of Econometrics,* **127**(1), 97–126.

Klein, L. R., and Ford, G. T. 2003. Consumer Search for Information in the Digital Age: An Empirical Study of Prepurchase Search for Automobiles. *Journal of Interactive Marketing,* **17**(3), 29-49.

Kleiner, M.M., and Kudrle, R.T. 2000. Does Regulation Affect Economic Outcomes? The Case of Dentistry. *Journal of Law and Economics,* 43(2), 547-582.

Koufaris, M., Kambil, A., and LaBarbera, P.A. 2001. Consumer Behavior in Web-Based Commerce: An Empirical Study, *International Journal of Electronic Commerce,* 6(2), 131-154.

Krishna, A., Wagner, M., Yoon, and Adaval, R. 2006. Effects of Effects of Extreme Priced Products on Consumer Reservation Prices. *Journal of Consumer Psychology,* 16(2), 176–190.

Krishnamurthi, L., Mazumdar, T., and Raj, S.P. 1992. Asymmetric Response to Price in Consumer Brand Choice and Purchase Quantity Decisions, *Journal of Consumer Research*, 19 (3), 387-400

Kuruzovich, J., Viswanathan, S., Agarwal, R., Gosain, S., and Weitzman, S. 2008. Marketspace or Marketplace? Online Information Search and Channel Outcomes in Auto Retailing. *Information Systems Research,* 19(2), 182-201.

Laurent, G. and Kapferer, J.N. 1985. Measuring Consumer Involvement Profiles, *Journal of Marketing Research,* 22(1), 41–53

Leonard, L., and Ariely, D. 2006. Shopping Goals, Goal Concreteness, and Conditional Coupons, *Journal of Consumer Research,* 33, 60-70.

Li, K., Kuo, C. and Russel, M.G. 1999. The Impact of Perceived Channel Utilities, Shopping Orientations, and Demographics on the Consumer's Online Buying Behavior, *Journal of Computer Mediated Communication,* 5 (2), online

Leland, H.E. 1981. The Informational Role of Warranties and Private Disclosure About Product Quality: Comment, *Journal of Law and Economics*, 24(3), 485-489

Lewis, G. 2009. Asymmetric Information, Adverse Selection and Online Disclosure: The Case of eBay Motors, *Harvard University Working Paper.*

Lizzeri, A. 1999. Information Revelation and Certification Intermediaries. *Rand Journal of Economics,* **30**(2), 214-231.

Lucas, P. 2001. How Product Demos are Helping Shoppers Kick the Tires on the Web, *Internet Retailer*, http://www.internetretailer.com/internet/marketing-conference/91973-how-product-demos-are-helping-shoppers-kick-tires-web.html (Feb)

Lynch, J. G., Jr., and Ariely, D. 2000. Wine Online: Search Cost and Competition on Price, Quality, and Distribution, *Marketing Science*, 19(Winter), 83–103.

Mack, D. 2009. Are People Shopping Online for Big-Ticket Items? Accessed at
http://blogs.adobe.com/scene7/2009/06/are_people_shopping_online_for.html

Maddala, G. S. 1987. Limited Dependent Variable Models Using Panel Data, *Journal of Human Resources* 22(3), 305-338.

Maddala, G. S. 1992. *Introduction to Econometrics*. McMillan, New York, NY.

Manchanda, P., Dube, J-P., Goh, K Y and Chintangunta, P.K. 2006. The Effects of Banner Advertising on Internet Purchasing, *Journal of Marketing Research*, **43** (February), 98-108.

Mandel, N., and Johnson, E.J. 2002. When web pages influence choice: Effects of visual primes on experts and novices, *Journal of Consumer Research*, 29, 235-245

Manheim. 2007. *Used Car Market Report*. 12th Edition

Mazar, N., B. Koszegi, and D. Ariely. 2009. Price Sensitive Preferences. *Berkeley Working Paper*. Available at elsa.berkeley.edu/~botond/PriceSensitivePreferences.pdf

McAfee Report. 2010. *Digital Window Shopping: The Long Journey to Buy*. Accessed at
http://www.mcafeesecure.com/pdf/Digital-Window-Shopping_FinaL-web.pdf

McLachlan, G. J., and Peel, D. 2000. *Finite mixture models*. New York: J. Wiley.

McWilliams, G. 2004. Analyzing Customers, Best Buy Decides Not All Are Welcome, *Wall Street Journal*. http://online.wsj.com/article/0,,SB109986994931767086,00.html (November 8)

Milgrom, P. 1981. Good News and Bad News: Representation Theorem and Applications, *Bell Journal of Economics*, 17(1), 380-391

Milgrom, P. 2008. What the Seller Won't Tell You: Persuasion and Disclosure in Markets, *Journal of Economic Perspectives*, 22(2), 115-131

Mitchell, J. 2008. *The Best and Worst Certified Cars.* Accessed online at
http://autos.aol.com/article/best-certified-used-cars/ (May 2)

Mitra, A. and J.G. Lynch, Jr. 1995. Towards a Reconciliation of Market Power and Information Theories of Advertising Effects on Price Elasticity, *Journal of Consumer Research*, 21(March), 644-659.

Moe, W. 2003. Buying, Searching, or Browsing: Differentiating between Online Shoppers using In-Store Navigational Clickstream, *Journal of Consumer Psychology*, 13(1&2), 29-40.

Moe, W., and Fader, P. 2004. Dynamic Conversion Behavior at E-Commerce Sites, *Management Science*, 50 (3), 326-335.

Montgomery, A., Hosanagar, K., Krishnan, R., and Clay, K. B. 2004. Designing a Better Shopbot, *Management Science*, 50(2), 189-206.

Murthi, B. P. S. and Sarkar, S. 2003. The Role of the Management Sciences in Research on Personalization, *Management Science*, 49(10), 1344-1362.

Novak, T., Hoffman, D. and Yung, Y. 2000. Measuring the customer experience in online environments: a structural modelling approach, *Marketing Science*, 19(1), 22–42.

Olshavsky, R.W. 1985. *Perceived quality in consumer decision-making: an integrated theoretical perspective*, DC Heath and Co., Lexington, MA.

Palmer, J. W. 2002. Web site usability, design, and performance metrics, Information Systems Research 13(2), 151–167.

Palmisano, C., Tuzhilin, A., and Gorgoglione, M. 2007. In B. N. Kokinov, D. C. Richardson, T. Roth-Berghofer, and L. Vieu, editors, *User Profiling with Hierarchical Context: An e-Retailer Case Study*, Lecture Notes in Computer Science, Springer, 369-383

Park, Y-H. and Fader, P.S. 2004. Modeling Browsing Behavior at Multiple Websites, *Marketing Science*, 23(3), 280-303.

Payne J.W., Bettman J.R., and Johnson E.J. 1993. *The adaptive decision maker*. Cambridge University Press: NY.

Perkowitz, M., and Etzioni, O. 1998. Adaptive Web Sites: Automatically Synthesizing Web Pages, *Proceedings of the 15th National Conference on Artificial Intelligence*, 727-732

Porter, H. and P. Sattler. 1999. Patterns of Trade in the Market for Used Durables: Theory and Evidence, *NBER Working Papers*.

Pratt, M.D., and Hoffer, G. E. 1985. State Used Motor Vehicle Disclosure Laws: Do They Make a Difference to the Consumer? *Journal of Consumer Affairs,* 19(1), 177-185.

Pratt, M.D. and Hoffer, G.E. 1986. The Efficacy of State Mandated Minimum Quality Certification: The Case of Used Vehicles, *Economic Inquiry*, 24(2), 313-318

PriceGrabber Consumer Behavior Report. 2009. Accessed at https://mr.pricegrabber.com/Online_Home_Furniture_Appliance_and_At_Home_Entertainment_Trends_May_2009_CBR_1.pdf

Purohit, D. 1992. Exploring the Relationship Between the Markets for New and Durable Goods: The Case of Automobiles, *Marketing Science*, 11(2), 154-167

Putsis, W.P., and Srinivasan, N. 1994. Buying or just browsing—The duration of purchase deliberation, *Journal of Marketing Research,* 31(3), 393–402.

Ratchford, B.T., Lee, M-S, and Talukdar, D. 2003. The Impact of the Internet on Information Search for Automobiles. *Journal of Marketing Research,* 40(May), 193-209.

Ratchford, B.T., Lee, M-S, and Talukdar, D. 2007. The Impact of the Internet on Information Search for Automobiles: A Re-Inquiry, *Journal of Consumer Research,* 34(1), 111-119

Ronnen, U. 1991. Minimum Quality Standards, Fixed Costs, and Competition, *RAND Journal of Economics*, 22, 490-504

Rook DW. 1987. The buying impulse, *Journal of Consumer Research,* 14, 189–99

Rook, D.W., and Gardner, M.P. 1993. In the mood: impulse buyings' affective antecedents, In: Janeen A-C, Russell WB, editors. *Research in consumer behavior*, 6, 1 – 28, Greenwich (CT): JAI Press.

Rosenbaum, P.R. 2002. Observational Studies. 2nd ed. New York, NY: Springer.

Rosenbaum, P., and Rubin, D. 1983. The Central Role of the Propensity Score in Observational Studies for Causal Effects, *Biometrika,* 70, 41–55.

Rossi, P.E., McCulloch, R.E. and Allenby, G.M. 1996. The Value of Purchase History Data in Target Marketing, *Marketing Science*, 15, 321-340.

Schadler, T., and Golvin, S. 2005. *The state of consumers and technology: Benchmark 2005.* Accessed at http://www.forrester.com/Research/Document/Excerpt/0,7211,36987,00.html

Schmidt, J.B. and Spreng, R.A. 1966. A Proposed Model of External Consumer Information Search, *Journal of the Academy of Marketing Science*, 24(3), 246-256.

Schwarz, G. 1978. Estimating the dimension of a model*, Annals of Statistics, 6*, 461-464.

Shapiro, C., and Varian, H.R. 1999. *Information Rules*. Boston, Mass.: Harvard Business School Press.

Shim, S., Eastlick, M.A. Lotz, S.L., and Washington, P. 2001. An Online Pre -purchase Intentions Model: The Role of Intention to Search, *Journal of Retailing*, 77, 397-416

Simonson, I. 1989. Choice Based on Reasons: The Case of Attraction and Compromise Effects, *Journal of Consumer Research*, 16(September), 158-174

Sismeiro, C., and Bucklin, R.E. 2004. Modeling Purchase Behavior at an E-Commerce Web Site: A Task Completion Approach, *Journal of Marketing Research*, 41,306-323

Sivakumar, K. and Raj, S.P. 1997. Quality tier competition: how price change influences brand choice and category choice, Journal of Marketing, 61(3), 71-84.

Spiliopoulou, M., Pohle, C., and Faulstich, L.C. 1999. Improving the effectiveness of a web site with web usage mining, In *Proceedings of KDD' 99 Workshop on Web Usage Analysis and User Profiling*, San Diego, CA, Aug. 1999. ACM (Masand and Spilipoulou eds.).

Srivastava, J., Cooley R., Deshpande, M., and Tan, P.N. 2000. Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data, *SIGKDD Explorations*, 1(2), 12-23.

Staiger, D., and Stock, J. H. 1997. Instrumental Variables Regression with Weak Instruments. *Econometrica*, 65(3), 557–86.

Stock, J. H., and Yogo, M. 2005. Testing for weak instruments in linear IV regressions. In D.W.K. Andrews and J.H. Stock (eds.), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*. Cambridge University Press, Cambridge.

Stock, J.H, Wright, J.H., and Yogo, M. 2002. A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments, *Journal of Business and Economic Statistics*, 20(4), 518-529

Stolyarov, D. 2002. Turnover of Used Durables in a Stationary Equilibrium: Are Older Goods Traded More? *Journal of Political Economy*, 110(6), 1390-1413

Sullivan, M.1998. How Brand Names Affect the Demand for Twin Automobiles, *Journal of Marketing Research*, 35(2), 154-165

Sultan, A. 2010. Can Leasing and CPO Reduce Adverse Selection in the Used Car Market? *Atlantic Economic Journal*, 38(1), 115-116

Tauscher, L., and Greenberg, S. 1997. How people revisit Web pages: Empirical findings and implications for the design of history systems, *International Journal of Human Computer Studies, Special issue on World Wide Web Usability* 47(1), 97-138.

Tedeschi,B. 2006. Buying a new fridge? Watch the interactive video first. nytimes.com/2006/05/15/technology/15ecom.html

Teicher,H . 1963. Identifiability of Finite Mixtures, *T he Annals of Mathematical Statistics*, 34, 1265-1269.

Terlaak, A., and King, A.A. 2006. The effect of certification with the ISO 9000 Quality Management Standard: A signaling approach. *Journal of Economics in Behavior and Organization,* 60(4), 579-602.

Thaler, R.H. 1999. Mental Accounting Matters, *Journal of Behavioral Decision Making*, 12(3), 183-206

Titterington, D., Smith, A., and Makov, U. 1985. Statistical Analysis of Finite Mixture Distributions, John Wiley and Sons.

Titus, P.A. and Everett, P.B. 1995. The consumer retail search process: a conceptual model and research agenda, *Journal of the Academy of Marketing Science*, 23(2), 106–119.

Trope, Y., and Liberman, N. 2003. Temporal Construal, *Psychological Review*, *110*, 403-421.

Urbany, J.E., Dickson, P.R., and Kalapurakal, R. 1996. Price Search in the Retail Grocery Market, *Journal of Marketing*, 60(2), 91–104.

Vella, M. 2006. *Best Used Cars.* Business Week Online (July 6).

Venkatesh, V., and Agarwal, R. 2006. From Visitors to Customers: A Usability-Centric Perspective on Purchase Behavior in Electronic Channels, *Management Science*, 52(3), 367-382.

Viscusi, K. 1978. A Note on Lemons Market with Quality Certification, *Bell Journal of Economics: The RAND Corporation*, 9(1), 277-279

Viswanathan, S., Kuruzovich, J.N., Gosain, S., and Agarwal, R. 2007. Online Infomediaries and Price Discrimination: Evidence from the Auto-Retailing Sector, *Journal of Marketing*, 71(3), 98-107.

Vriens, M., Wedel, M. and Wilms, T. 1996. Metric Conjoint Segmentation Methods: A Monte Carlo Comparison, *Journal of Marketing Research*, 33(February), 73–85.

Wedel, M., and Kamakura, W.A. 2000. *Market Segmentation: Conceptual and Methodological Foundations*. Dordrecht: Kluwer.

Wedel M., and Wagner, K. 2002. Introduction to the special issue on market segmentation, *International Journal of Research in Marketing*, 19 (3), 181-183.

Wimmer, B.S., and Chezum, B. 2003. An Empirical Examination of Quality Certification in a Lemons Market, *Economic Inquiry,* 41(2), 279-291.

Wernerfelt, B. 1995. A Rational Reconstruction of the Compromise Effect: Using Market Data to Infer Utilities, *Journal of Consumer Research*, 21(March), 627-633.

Wetzel, J. and G. Hoffer. 1982. Consumer Demand for Automobiles: A Disaggregated Market Approach, Journal of Consumer Research, 9(2), 195-199

Winer, R.S., Deighton, J., Gupta, S., Johnson, E.J., Mellers, B., Morwitz, V.G., O'Guinn, T., Rangaswamy, A. and Sawyer, A.G. 1997. Choice in Computer-Mediated Environments", *Marketing Letters,* 8(3), 287–96.

Wolfinbarger, M. and Gilly, M. 2001. Shopping online for freedom, control, and fun, California Management Review, 43(2), 35–55.

Wooldridge, J. 2002. *Econometric Analysis of Cross Section and Panel Data*, MIT Press.

Wu, J., Cook, V.J. Jr., and Strong, E.C. 2005. A Two-Stage Model of the Promotional Performance of Pure Online Firms, *Information Systems Research*, 16(4), 334-351

Yang, Y. and Padmanabhan, B. 2007. Behavioral Signatures in Web Browsing Data, *In Proceedings of the First China Summer Workshop on Information Management,* July 22-23

Yankee Group and Ernst and Young. 2002. What Causes Customers to Buy on Impulse?, *User Interface Engineering,* http://www.uie.com/publications/whitepapers/ImpulseBuying.pdf

Zeithaml, V. 1988. Consumer Perceptions of Price, Quality and Value: A Means-end Model and Synthesis of Evidence. *Journal of Marketing,* 52(3), 2-22.

Zettelmeyer, F., Scott Morton, F., and Silva-Risso, J. 2006. How the Internet Lowers Prices: Evidence from Matched Survey and Auto Transaction Data, Journal of Marketing Research, 43(2), 168-181.

Zettelmeyer, F., Scott Morton, F., and Silva-Risso, J. 2005. Cowboys or Cowards: Why Are Internet Car Prices Lower? *Discussion paper, Haas School of Business, University of California, Berkeley*.

Zhang, J. and Krishnamurthi, L. 2004. Customizing promotions in online stores, *Marketing Science* 23(4), 561-578.