

Marketing Science

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

Simultaneous or Sequential? Search Strategies in the U.S. Auto Insurance Industry

Elisabeth Honka, Pradeep Chintagunta

To cite this article:

Elisabeth Honka, Pradeep Chintagunta (2017) Simultaneous or Sequential? Search Strategies in the U.S. Auto Insurance Industry. Marketing Science 36(1):21-42. <https://doi.org/10.1287/mksc.2016.0995>

Full terms and conditions of use: <http://pubsonline.informs.org/page/terms-and-conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2017, INFORMS

Please scroll down for article—it is on subsequent pages

INFORMS is the largest professional society in the world for professionals in the fields of operations research, management science, and analytics.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

Simultaneous or Sequential? Search Strategies in the U.S. Auto Insurance Industry

Elisabeth Honka,^a Pradeep Chintagunta^b

^a University of California Los Angeles, Los Angeles, California 90024; ^b Booth School of Business, University of Chicago, Chicago, Illinois 60637

Contact: elisabeth.honka@anderson.ucla.edu (EH); pradeep.chintagunta@chicagobooth.edu (PC)

Received: April 29, 2014

Accepted: February 15, 2016

Published Online in Articles in Advance:
August 1, 2016

<https://doi.org/10.1287/mksc.2016.0995>

Copyright: © 2017 INFORMS

Abstract. We study the identification of the search method consumers use when resolving uncertainty in the prices of alternatives. We show that the search method—simultaneous or sequential—is identified with data on consumers’ consideration sets (but not the sequence of searches), prices for the considered alternatives and marketwide price distributions. We show that identification comes from differences in the patterns of actual prices in consumers’ consideration sets across search methods. We also provide a new estimation approach for the sequential search model that uses such data. Using data on consumer shopping behavior in the U.S. auto insurance industry that contain information on consideration sets and choices, we find that the pattern of actual prices in consumers’ consideration sets is consistent with consumers searching simultaneously. Via counterfactuals we show that the consideration set and purchase market shares of the largest insurance companies are overpredicted under the incorrect assumption of sequential search. As the search method affects consumers’ consideration sets, which in turn influence brand choices, understanding the nature of consumer search and its implications for consideration and choice is important from a managerial perspective.

History: Russell Winer served as the senior editor and Chakravarthi Narasimhan served as associate editor for this article.

Supplemental Material: Data and the online appendix are available at <https://doi.org/10.1287/mksc.2016.0995>.

Keywords: consumer search • simultaneous search • sequential search • auto insurance industry

1. Introduction

Understanding the formation of consideration sets and their subsequent implications for consumers’ product choices has long been an area of interest to marketers (e.g., Hauser and Wernerfelt 1990). Accordingly, researchers have tried to understand how these sets are formed (Hauser and Wernerfelt 1990) or have tried to account for them when studying choice behavior (Siddarth et al. 1995, Chiang et al. 1999, Mehta et al. 2003, Seiler 2013). In the latter case, even in the absence of data on consideration sets, researchers have tried to incorporate the notion of consideration sets via functional form assumptions on consideration set and choice probabilities. Explicitly accounting for the role of consideration sets in consumers’ decision-making is important from the perspective of correctly measuring consumers’ brand preferences and their sensitivities to marketing activities, as failure to do so could lead to incorrect inferences regarding these market fundamentals. At the same time, research in economics has shown that the process by which a consumer arrives at his consideration set also has implications for these parameters and consequently for firms operating in the market.

The theoretical underpinnings of consideration set formation are in the models of search.¹ A consumer who engages in search is uncertain about some dimension(s) of the product or service, say, price, and resolves this uncertainty by incurring a search cost. In the search process, the consumer trades off the costs incurred and benefits accrued from the undertaking to arrive at a consideration set for which he has complete information. At this stage, the consumer is back to the familiar choice situation of complete information that has been extensively studied in the marketing literature (e.g., the brand choice literature using scanner panel data). If a consumer incurs a marginal cost for each product or service searched, then the number of options the consumer ends up considering before making a choice critically depends on the search strategy the consumer uses.²

In this paper, we look at situations in which consumers are uncertain about price (but not the other attributes of the product) and engage in costly search to resolve this price uncertainty. We study two search methods, namely, simultaneous and sequential search. Under a simultaneous search strategy, the consumer samples a fixed number of alternatives and purchases the alternative with the lowest price (or

highest utility) in this set. The set of alternatives searched is obtained by looking at the subset for which the expected maximum utility net of search costs is the highest among all possible subsets. A limitation of the simultaneous search strategy is that it does not take into account new information that the consumer might obtain during the search process. So if the consumer observes a very low price (or very high utility) for an alternative early in the search process, the benefit from an additional search may be below the marginal cost of that search. In a sequential search strategy, on the other hand, the number of alternatives searched is not fixed, but is a random variable that depends on the outcome of the search; this allows a consumer to economize on information costs. In this case, the consumer weighs the expected benefits and costs of gathering additional price information after each new quote is obtained. If an acceptable price is obtained early on, the expected gains from additional searches are small and there is no need to pay the cost of additional searches (see Baye et al. 2006).

Since in most instances researchers only observe variation in prices or purchase outcomes, it is not possible to identify the search method with just these data. Previous empirical research has circumvented this challenge by explicitly assuming the type of search that consumers engage in. For example, Mehta et al. (2003), Pires (2015), and Muir et al. (2013) assume that consumers search simultaneously, while Dahlby and West (1986), Kim et al. (2010), and Chen and Yao (2016) assume that consumers search sequentially. In this paper, we focus on the case where consumers engage in price search and the researcher observes each consumer's consideration set (but not the sequence of searches), besides purchase outcomes, prices, price distributions, and other characteristics. We show that, under certain assumptions, the search method is indeed identified by the price patterns in consumers' observed consideration sets. Differences in price patterns emerge because in simultaneous search consumers only use information on the expected prices to decide which and how many companies to search, whereas in sequential search consumers continue searching and consider more alternatives only when they receive high price draws for the initially considered alternatives. Our identification strategy holds for a broad range of settings that we discuss in detail in Section 3.2.

Next, we examine the consequences of imposing an incorrect search method assumption on the estimated consumer preference and search cost parameters when researchers have access to the above data. To accomplish this, we first need an estimation approach for the sequential search model where the researcher has access to individual-level data on consideration sets, purchases, and other characteristics, but not the

sequence of searches. We avoid having to enumerate all possible search sequences by placing a small set of restrictions on consumers' utilities and reservation utilities. These restrictions are derived from Weitzman's (1979) selection, stopping, and choice rules and the insight that, in addition to Weitzman's (1979) rules, it must have been optimal for the consumer not to stop searching and purchase earlier. Similar to the simultaneous search model for which we apply a simulated maximum likelihood estimation (SMLE) approach suggested by Honka (2014), we propose an SMLE-based approach for the sequential search model. Using extensive simulations we are able to show that incorrect assumptions on the search method could lead to different consideration sets and biased estimates of preference parameters and search costs.

We provide an empirical application of our search method identification strategy and new sequential search estimation approach in the context of the U.S. auto insurance industry. Using data on consumers' consideration sets, purchases, prices, and other characteristics, we first ask: do households search simultaneously or sequentially when shopping for auto insurance price quotes? Since consumers in our sample have been insured previously and coverage levels tend not to change much, assuming that consumers engage in price search is a reasonable assumption in this context. We look for model-free evidence of a search method taking the auto insurance industry-specific practice of sending customers a renewal offer into account and then estimate the model parameters under the assumptions of simultaneous and sequential search. We find both the model-free evidence and the estimates to provide support for simultaneous search. Our estimated search cost is \$42. We then study via counterfactuals how elasticities and market shares are influenced by an incorrect assumption on the search method. We find that consideration set and purchase market shares of the largest four insurance companies are overpredicted under the incorrect assumption of sequential search. We then assess the robustness of our results to the presence of unobserved heterogeneity in the search method and to assumptions required by our estimation methods.

The main contributions of this paper are as follows. First, we show both analytically and in simulations that the search method consumers use is identified by the price patterns in consumers' consideration sets for a very broad range of settings. Second, we provide a comparison of the consequences of assuming simultaneous versus sequential search strategies on the parameter estimates in contexts where the only data available to researchers besides typically available choice data are information on consumers' consideration set compositions. These kinds of data are becoming more widely available across a variety of

service businesses as well as from surveys conducted by firms such as J.D. Power for a variety of categories (e.g., automobile purchases, hotels, and retail banking). Third, in providing such a comparison, we need to be able to estimate model parameters under both search method assumptions for these kinds of data. While Honka (2014) provides an approach for simultaneous search that we adopt here, we propose an estimation approach under the sequential search assumption. Finally, we quantify the effects of assuming the incorrect search method on quantities that are typically of interest to researchers such as elasticities and market shares.

In the next section, we discuss the relevant literature. In Section 3, we introduce our model and discuss search method identification. In Section 4, we describe our estimation approaches and summarize results from Monte Carlo studies. In Section 5, we discuss our empirical application, and in Section 6, we study a counterfactual. In Section 7, we check the robustness of our results. We close our paper by discussing its limitations and future research opportunities and finally conclude.

2. Relevant Literature

The topic of consideration sets has seen much interest in the marketing literature as it sits at the intersection of economics (starting with the work of Stigler 1961, who looks at the costs and benefits of gathering information on brands) and psychology (e.g., Miller 1956 notes the cognitive challenges of processing information on all of the brands). In line with this basic idea, several studies in marketing have approached the area by looking at the costs and benefits associated with gathering information on brands (e.g., Ratchford 1980, Shugan 1980, Hauser and Wernerfelt 1990, Roberts and Lattin 1991, Mehta et al. 2003, etc.). Many aspects of consideration sets have been investigated in the literature, including the benefits to the prediction of choice (e.g., Siddarth et al. 1995, Andrews and Srinivasan 1995, Bronnenberg and Vanhonacker 1996), greater diagnostic insight into the choice process (e.g., Gensch and Soofi 1995), and understanding the antecedents of consideration (DeSarbo and Jedidi 1995, Mitra and Lynch 1996, DeSarbo et al. 1996). Roberts and Lattin (1997), surveying the marketing literature on consideration sets, highlight several potential areas of future research, including understanding the dynamics of consideration, similarity and differences in alternatives included in the set, and other behavioral dimensions of this phenomenon. Importantly, the authors also note insightfully that since authors of several of the studies in the literature use only data on consumers' final choices, i.e., "take no explicit measures of consideration, they cannot address whether the consideration stage of their model corresponds to a cognitive stage of

consideration in the consumer's decision process, or if it is just a statistical artifact of the data" (Roberts and Lattin 1997, p. 407). In this paper, we try to understand consumers' consideration and choice by looking at data from both of these stages in the choice process. Specifically, using the literature from economics on search as the underpinning of our empirical analysis, we investigate the identification of the search method as well as the estimation of simultaneous and sequential models of search behavior.

Our paper is embedded in the literature on consumer search. As this literature is extensive, we focus only on recent efforts to structurally estimate search models or to identify the search method from data. De los Santos et al. (2012) show that with data on purchases, consideration sets, and the sequence of searches, the search method consumers use is identified for both homogeneous and differentiated goods. In this paper, we focus on the situation where the researcher observes only consideration sets and purchases, but not the sequence of searches, i.e., the researcher has less information, and show that even in this case the search method is identified. Hong and Shum (2006) develop methodologies to estimate search costs under both simultaneous and sequential search when only prices are observed. In this paper, we are able to relax several of Hong and Shum's (2006) assumptions: For example, we allow goods to be differentiated and price distributions to be company-specific. Similar to Hong and Shum (2006), we are able to compare search costs under the two assumptions on search strategies. In a follow-up paper to Hong and Shum (2006), Chen et al. (2007) develop nonparametric likelihood ratio model selection tests that allow them to test between simultaneous and sequential search models. Chen et al. (2007) do not find significant differences between the simultaneous and sequential search models using the usual significance levels in their empirical application. Finally, this paper is also related to Honka (2014). She quantifies search and switching costs for the U.S. auto insurance industry using the same data we do in this paper. The simultaneous search model presented here is similar to the one Honka (2014) uses.

3. Model and Search Method Identification

3.1. Model

We present a general differentiated goods price search model: Consumers know the distribution(s) of prices in the market, but have to engage in costly search to learn the specific price a company is going to charge them. Given their search sets, consumers maximize the utility of the purchased option. As standard in the price search literature (see also Hong and Shum 2006, De los Santos et al. 2012, Honka 2014), we make the following set of assumptions:³ (a) prices are the only source of uncertainty for the consumer that he

resolves through search; (b) consumers know the distribution of prices and have rational expectations for these prices; (c) price draws are independent across companies; (d) there is no learning about the price distribution from observing other variables (e.g., advertising); (e) (search costs are sufficiently low so that) all consumers search at least once; and (f) consumers have nonzero search costs. Furthermore, our model allows for observed and unobserved heterogeneity in preferences and search costs.⁴

More specifically, there are N consumers, indexed by $i = 1, \dots, N$, who purchase one of J brands, indexed by $j = 1, \dots, J$. Consumer i 's indirect utility for company j is given by

$$u_{ij} = \alpha_{ij} + \beta p_{ij} + X_{ij}\gamma + \epsilon_{ij}, \quad (1)$$

where ϵ_{ij} are independently and identically distributed (i.i.d.) and observed by the consumer but not by the researcher, α_{ij} are brand intercepts, and p_{ij} are prices, which follow some well-defined distribution with mean μ_{ij}^p .⁵ Without loss of generality, we assume that $\beta < 0$. The term X_{ij} may contain other variables that influence consumer utility. These variables can be consumer-specific (e.g., demographics), company-specific (e.g., advertising spending), or both. The parameters to be estimated are α_{ij} , β , and γ .⁶

3.1.1. Simultaneous Search. To decide on the set of companies S_i to obtain prices for, the consumer calculates the net benefit of all possible search sets in terms of their size and composition. A consumer's net benefit of a searched set Γ_{S_i} is given by the expected maximum utility among the searched brands minus the cost of search

$$\Gamma_{S_i} = E \left[\max_{j \in S_i} u_{ij} \right] - k \cdot c_{ij}. \quad (2)$$

Once a consumer has formed his consideration set and learned the prices, all price uncertainty is resolved for this set. Both the consumer and the researcher observe prices. The consumer then picks the company with the highest utility among the searched companies, i.e.,

$$m = \arg \max_{j \in S_i} u_{ij}, \quad (3)$$

where u_{ij} now includes the quoted prices for consumer i by company j .

3.1.2. Sequential Search. We present a sequential search model with recall. Weitzman (1979) showed that it is optimal for a consumer to rank all companies according to their reservation utilities in decreasing order when deciding on the search sequence (selection rule). Reservation utility r_{ij} is the utility that makes a consumer indifferent between searching and not searching

$$c_{ij} = \int_{r_{ij}}^{\infty} (u_{ij} - r_{ij}) f(u_{ij}) du_{ij}. \quad (4)$$

A consumer stops searching when the maximum utility among the searched companies is larger than the maximum reservation utility among the nonsearched companies (stopping rule), i.e.,

$$\max_{j \in S_i} u_{ij} > \max_{j' \notin S_i} r_{ij'}. \quad (5)$$

Finally, the choice rule states that the consumer picks the company with the largest utility among the searched ones

$$m = \arg \max_{j \in S_i} u_{ij}. \quad (6)$$

Thus, after receiving each price draw, the consumer decides to either continue searching or to stop searching and purchase from the set of searched companies. Note that, in contrast to the simultaneous search model, the consideration and purchase stages are not separate.

3.2. Search Method Identification

We first discuss the price patterns we would observe under simultaneous and sequential search and then describe circumstances under which these patterns are distinct and those under which they are not; the latter provides conditions under which the search method is not identified.

We start out by discussing the data pattern that characterizes the simultaneous search method. Recall that prices follow some (potentially company-specific and/or consumer-specific) distribution(s). Let us define $\Pr(p < \mu_{ij}^p) = \lambda$, i.e., the probability that a price draw is below the expected price is λ . Furthermore, we define event $X = 1$ as a below-price-expectation price draw and $X = 0$ as an above-price-expectation price draw. Recall that under simultaneous search the search rule says that the consumer precommits to a search set S_i consisting of k_i companies. Then we can calculate the expected proportion of below-price-expectation prices in a consumer's consideration set of size k_i as

$$E \left[\frac{1}{k_i} \sum_{m=1}^{k_i} X_m \right] = \frac{1}{k_i} \sum_{m=1}^{k_i} E[X_m] = \frac{\lambda k_i}{k_i} = \lambda.$$

Thus, we expect λ percent of the price draws in consumers' consideration sets to be below and $1 - \lambda$ percent to be above the expected price(s). The crucial ingredients for identification are that the researcher observes the means of the price distributions μ_{ij}^p , the actual prices in consumers' consideration sets p_{ij} , and the probability of a price draw being below its mean λ .

We now turn to sequential search and the data pattern that is characteristic for this search method. We present here the proof for differentiated goods with consumer- and company-specific search costs and refer the reader to Online Appendix B for detailed analytical proofs for both homogeneous and differentiated

goods. Recall that consumers have a utility function as described in Equation (1), with $\beta < 0$ and $\epsilon_{ij} \sim \text{iid}$. Consumers have search costs $c_{ij} > 0$ and make k_i searches. The probability of getting a below-price-expectation price draw is λ , i.e., $P(p < \mu_{ij}^p) = \lambda$. Given the assumptions for the price distributions, consumers' utility also follows some well-defined distribution with mean $\mu_{ij} = \alpha_{ij} + \beta \mu_{ij}^p + X_{ij} \gamma + \epsilon_{ij}$. Furthermore, since $P(p < \mu_{ij}^p) = \lambda$, it must be that $P(u > \mu_{ij}) = \lambda$, i.e., a below-price-expectation price draw always results in an above-mean level of utility.

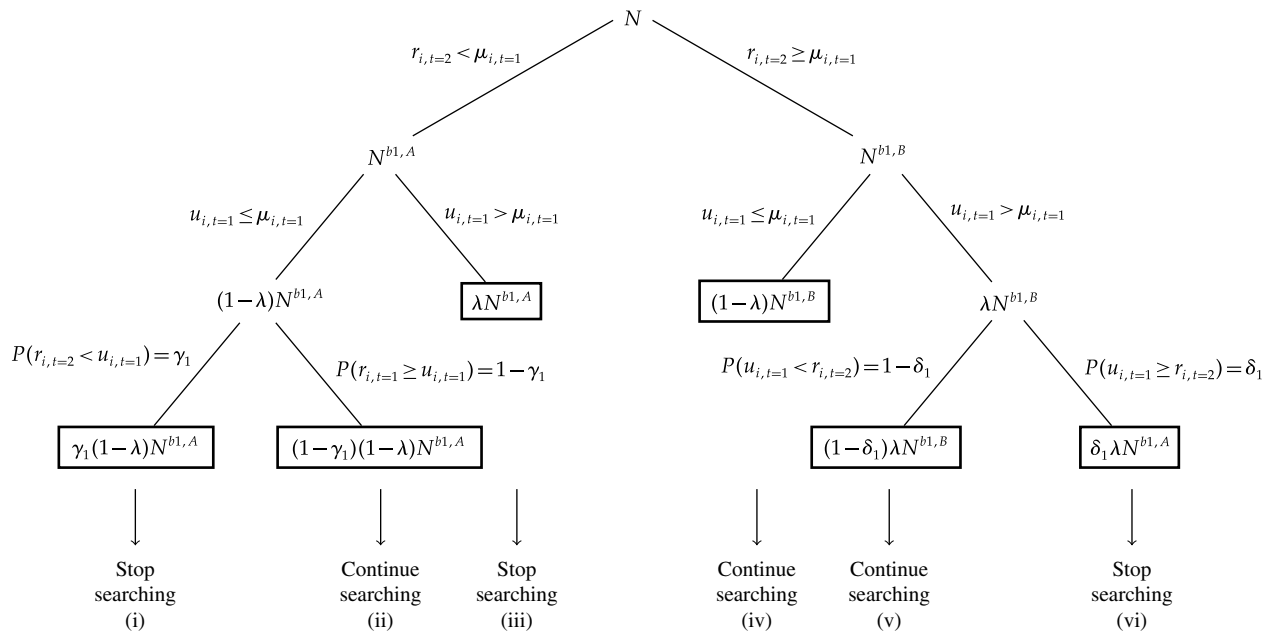
Using Weitzman's (1979) selection rule, we know that consumers order all alternatives in a decreasing order of their reservation utilities r_{ij} . Consumers first search the alternative with the highest reservation utility, then the alternative with the second highest reservation utility, etc. To express the ranking according to the reservation utilities r_{ij} , let us define $r_{i,t=1}$ as the company with the highest reservation utility for consumer i , $r_{i,t=2}$ as the company with the second highest reservation utility for consumer i , etc. Using Weitzman's (1979) stopping rule, we know that consumers stop searching when the maximum utility among the searched alternatives is larger than the maximum reservation utility among the nonsearched companies. In the following, we characterize the proportion of consumers who receive a price draw below the expected price for the brands in their consideration sets when these sets are of size one.

Before the first search (denoted by "b1") we classify consumers into one of two types—Type A and Type B—with $N = N^{b1,A} + N^{b1,B}$.⁷ Type A (B) consumers are those whose reservation utility of the

potentially second-to-be-searched company is smaller (larger) than the expected utility of the company searched first, i.e., $r_{i,t=2} < \mu_{i,t=1}$ ($r_{i,t=2} \geq \mu_{i,t=1}$). Note that according to Weitzman's (1979) rule, consumers can change their type, i.e., change from being Type B to being Type A (but not vice versa!) as they move from the first to the second to the third search, etc. In Figure 1, we characterize what happens to these consumers after their first search.

Consumers who stop searching after the first search include the following: (a) Type A consumers whose realized utility in the first search exceeds the expected utility of the company searched first and therefore, by the definition of Type A consumers, is also larger than the reservation utility of the company potentially to be searched second (a fraction λ of Type A consumers); (b) Type A consumers whose price draw is such that the utility of the first search at the realized price draw is below the expected utility of the first search (the remaining fraction $(1 - \lambda)$) and whose reservation utility of the company potentially to be searched second is nevertheless below the realized utility level of the company searched first (a further fraction $0 \leq \gamma_1 \leq 1$). This group represents a fraction $(1 - \lambda)\gamma_1$ of Type A consumers; and (c) Type B consumers who get a low price draw and hence whose realized utility in the first search is larger than the expected utility of the company searched first (a fraction λ) and whose reservation utility of the company potentially to be searched second is below this realized level of utility of the company searched first (a further fraction $0 < \delta_1 \leq 1$). This group represents a fraction $\lambda\delta_1$ of Type B consumers.

Figure 1. Consumers' Decision Making During First Search Under Sequential Search



Thus, the number of consumers who stop searching after the first search is $N_1 = \lambda N^{b1,A} + (1 - \lambda)\gamma_1 N^{b1,A} + \lambda \delta_1 N^{b1,B}$. Of these consumers, those in (a) and (c) are the ones that get a below mean price draw. From this we can compute the fraction of consumers that receive a below mean price draw among those who only search once as

$$X_1 = \frac{\lambda N^{b1,A} + \lambda \delta_1 N^{b1,B}}{\lambda N^{b1,A} + (1 - \lambda)\gamma_1 N^{b1,A} + \lambda \delta_1 N^{b1,B}}. \quad (7)$$

In Online Appendix B, we show that this proportion X_1 is always larger than λ for differentiated goods with consumer- and company-specific search costs under the necessary condition that we observe a positive number of consumers in the data making more than one search.

Suppose in the data we observe consumers making $k = 1, \dots, K$ searches with $K > 1$, i.e., a setting in which the search method is identified because we cannot get the same price pattern under both simultaneous and sequential search. We showed that, under simultaneous search, the proportion of below-price-expectation price draws in consumers' consideration sets is constant and equals λ for all k . Under sequential search, the proportion of below-price-expectation price draws in consumers' consideration sets of size one is always larger than λ . Thus, with individual-level data on consumers' consideration sets on hand, the researcher needs to look at the pattern of below-price-expectation price draws among consumers searching once: If that proportion equals λ , consumers are searching simultaneously. If that proportion is larger than λ , consumers are searching sequentially.

3.2.1. Discussion. The characteristic price patterns for simultaneous and sequential search described above hold for all models that satisfy the assumptions stated at the beginning of Section 3.1. This includes (1) models for homogeneous goods, (2) models for differentiated products, (3) models that include unobserved heterogeneity in preferences and/or search costs, (4) models with correlations among preferences and search costs, and (5) models with observed heterogeneity in price distribution means μ_{ij} . On the other hand, we do not find the characteristic price patterns when there is unobserved heterogeneity in the price distribution means as the researcher would no longer be able to judge whether a price draw is above or below the mean. Note also that our identification arguments here are based on the first moments of prices; in principle, there could be identification rules based on higher moments as well. We leave this investigation for future research.

Next, we discuss the modeling assumptions stated at the beginning of Section 3.1 and to what extent our search method identification results depend on them. Recall that the assumptions are as follows: (a) prices are

the only source of uncertainty for the consumer that he resolves through search; (b) consumers know the distribution of prices and have rational expectations for these prices; (c) price draws are independent across companies; (d) there is no learning about the price distribution from observing other variables (e.g., advertising); (e) (search costs are sufficiently low so that) all consumers search at least once; and (f) consumers have nonzero search costs.

Assumptions (a) through (d) are standard in both the theoretical and empirical literature on price search, e.g., Stigler (1961), Weitzman (1979), Morgan and Manning (1985), Mehta et al. (2003), Hong and Shum (2006), Chen et al. (2007), Moraga-Gonzalez and Wildenbeest (2008), De los Santos et al. (2012), Honka (2014), and Honka et al. (2016). With regard to the rational expectations part of assumption (b), note that—just like almost all previous literature on consumer search (barring a few exceptions⁸)—we rely on the rational expectations assumption to prove search method identification. If (the researcher believes that) consumers do not have rational price expectations or these rational price expectations cannot be estimated from data, our result for the characteristic price pattern under simultaneous search still holds. However, we are no longer able to predict the characteristic pattern under sequential search.⁹ Thus, the researcher is able to (visually or statistically¹⁰) test the (stand-alone) assumption of simultaneous search. If he is able to reject the null hypothesis of simultaneous search, it means that consumers do not search simultaneously. However, not being able to reject the null hypothesis of simultaneous search does not imply that the researcher can say anything about consumers (not) searching sequentially; i.e., the results from any hypothesis test about simultaneous search have no implications for whether consumers search/do not search sequentially when consumers have irrational price expectations.

With regard to assumption (e) that (search costs are sufficiently low so that) all consumers search at least once, three approaches are common in the search literature to deal with consumers' search/no-search decisions: Researchers have assumed that (i) (search costs are sufficiently low so that) all consumers search at least once (e.g., Reinganum 1979), (ii) the first search is free (e.g., Stiglitz 1987, Stahl 1989), and (iii) all consumers use the same search method, but it is optimal for some consumers to search/not to search depending on the level of their search costs (e.g., Janssen et al. 2005).¹¹ Throughout this paper, we use assumption (i) for expository purposes. However, all of our results hold across the three possible assumptions.¹²

With regard to assumption (f) that consumers have nonzero search costs, note that search costs have to be only *marginally* larger than zero for search

method identification to hold in all model specifications (see Online Appendix B). When search costs are exactly zero for some consumers or some companies, our search method identification results hold for *all* homogeneous and differentiated goods models with constant (across consumers and companies), consumer-specific, or company-specific search costs. In this case, the necessary condition is that the distribution of consideration set sizes has to have at least two mass points (this is a straightforward generalization of the necessary condition that a positive number of consumers has to search more than once). Additionally, a sufficient but not necessary condition is that consumers do not only search one or all companies in the market. We show this in detail in Online Appendix C. *Only* to achieve search method identification for (homogeneous or differentiated goods) models with unobserved heterogeneity in search costs across *both* consumers *and* companies, we need the assumption that all consumers have nonzero search costs.¹³ Alternatively, if the researcher believes that the assumption of nonzero search costs is not appropriate in an empirical setting, search method identification is also given under the assumption that the search cost distribution is continuous, i.e., has support, from 0 to a positive number $A > 0$.¹⁴ Again, this is only needed for models with unobserved heterogeneity in search costs across both consumers and companies. We provide details on search method identification when some search costs are exactly zero in Online Appendix C.

4. Estimation

We now present estimation approaches for the simultaneous and sequential search models as described in Section 3.1. Our goal is to present estimation approaches that can be used for markets with *any* number of alternatives. This is a challenge for the estimation of a simultaneous search model that suffers from the curse of dimensionality (Chiang et al. 1999, Kim et al. 2010). To overcome this challenge, we use the theory developed by Chade and Smith (2005). Their theory can only be used under two conditions: (1) there is first-order stochastic dominance among the price distributions and (2) search costs cannot be company-specific. We implement the first condition by assuming that the variances of the price distributions are identical.¹⁵ Note that these two assumptions are not necessary for the sequential search model, but that we nevertheless make them to keep everything other than the search method consistent across the simultaneous and sequential search models. Furthermore, our search method identification results do not rely on these two assumptions as described in Section 3.2.¹⁶

We start by pointing out the crucial differences between what the consumer observes and what the researcher observes: First, while the consumer knows

the distributions of prices in the market, the researcher does not. Second, while the consumer knows the sequence of searches, the researcher only partially observes the sequence by observing which companies are being searched and which ones are not being searched. Third, in contrast to the consumer, the researcher does not observe ϵ_{ij} . To address the first issue that the researcher does not observe the price distributions, these distributions need to be inferred from the data. In other words, the typical assumption of rational expectations (e.g., Mehta et al. 2003, Hong and Shum 2006, Moraga-Gonzalez and Wildenbeest 2008) is that these distributions can be estimated from the prices observed in the data. However, since the parameters of the distributions thus obtained are estimates, the associated sampling error needs to be accounted for when estimating the other parameters of the model (see McFadden 1986).

4.1. Simultaneous Search

The estimation approach for the simultaneous search model we present in this section is closely related to the one developed in Honka (2014). The main difference between the two models is that Honka (2014) assumes that prices follow an Extreme Value (EV) type I distribution, while we assume that prices follow a normal distribution with $p \sim N(\mu_{ij}^p, \sigma^p)$. This change in distributional assumption is driven by the desire to have the same distributional assumption on prices under both simultaneous and sequential search.¹⁷ Given the normal assumption for prices, the utility u_{ij} is a normally distributed random variable with mean $\mu_{ij} = \alpha_{ij} + \beta\mu_{ij}^p + X_{ij}\gamma + \epsilon_{ij}$ and standard deviation $\sigma = \beta\sigma^p$. A consumer's search decision under simultaneous search depends on the expected indirect utilities (EIUs) (Chade and Smith 2005). Consumer i 's EIUs where the expectation is taken with respect to price are given by

$$E[u_{ij}] = \alpha_{ij} + \beta E[p_{ij}] + X_{ij}\gamma + \epsilon_{ij}. \quad (8)$$

Consumer i observes these EIUs for every brand in his market (including ϵ_{ij}). To decide which companies to search, consumer i ranks all companies according to their EIUs (Chade and Smith 2005) and then picks the top k companies to search. The term O_{ik} denotes the set of top k companies consumer i ranked highest according to their EIUs. For example, O_{i1} contains the company with the highest expected utility for consumer i , O_{i2} contains the companies with the two highest expected utilities for consumer i , etc. To decide on the number of companies k to obtain prices for, the consumer calculates the net benefit of all possible search sets *given the ranking of EIUs*; i.e., if there are J companies in the market, the consumer can choose among J choice sets. A consumer's benefit of a searched set S_i is given by the expected *maximum* utility among the

searched brands. The consumer picks the size of his searched set S_i that maximizes his net benefit of searching, denoted by Γ_{ik} , i.e., expected maximum utility among the searched companies minus the cost of search

$$\Gamma_{ik} = E \left[\max_{j \in O_{ik}} u_{ij} \right] - k \cdot c. \quad (9)$$

Recall that the researcher does not observe the exact ranking according to the EIUs, but he observes the sets of companies that the consumer considers and does not consider. Honka (2014) has shown that this allows the researcher to describe a consumer's search set using the following two conditions:

$$\min_{j \in S_i} (E[u_{ij}]) \geq \max_{j' \notin S_i} (E[u_{ij'}]) \cap \Gamma_{ik} \geq \Gamma_{ik'} \quad \forall k \neq k'; \quad (10)$$

i.e., the minimum EIU among the searched brands is larger than the maximum EIU among the nonsearched brands *and* the net benefit of the chosen searched set of size k is larger than the net benefit of any other search set of size k' .

We account for the fact that the researcher does not observe ϵ_{ij} by assuming that ϵ_{ij} has an EV type I distribution with location parameter 0 and scale parameter 1 and integrate over its distribution to obtain the corresponding probabilities with which we can compute the likelihood function. Then the probability that a consumer picks a consideration set Υ is given by

$$P_{i\Upsilon|\epsilon} = P \left(\min_{j \in S_i} (E[u_{ij}]) \geq \max_{j' \notin S_i} (E[u_{ij'}]) \cap \Gamma_{ik} \geq \Gamma_{ik'} \quad \forall k \neq k' \right). \quad (11)$$

Let us now turn to the purchase decision given consideration. Let J be the base brand for consumer i . Then the consumer's choice probability conditional on his consideration set is

$$P_{ij|\Upsilon, \epsilon} = P(u_{ij} \geq u_{ij'} \quad \forall j \neq j', j, j' \in S_i). \quad (12)$$

Note that there is a selection issue: Given a consumer's search decision, the ϵ_{ij} do not follow an EV type I distribution, and the conditional choice probabilities do not have a logit form.

In summary, the researcher estimates the price distributions, only partially observes the utility rankings, and does not observe ϵ_{ij} in the consumer's utility function. Accounting for these differences compared to the consumer, we derived an estimable model with the consideration set probability given by Equation (11) and the conditional purchase probability given by Equation (12). We maximize the joint likelihood of consideration set and purchase. The likelihood of our model is given by

$$L = \prod_{i=1}^N \int_{-\infty}^{+\infty} \left(\prod_{l=1}^L \prod_{j=1}^J P_{i\Upsilon|\epsilon}^{\delta_{il}} P_{ij|\Upsilon, \epsilon}^{\delta_{ij}} \right) f(\epsilon) d\epsilon, \quad (13)$$

where δ_{il} indicates the chosen consideration set and δ_{ij} the company from which insurance is purchased. Neither the consideration set nor the conditional purchase probability have a closed-form solution. Honka (2014) describes how to estimate the simultaneous search model under the assumption of EV type I distributed prices in four steps in detail. Since our assumption of normally distributed prices results in no closed-form solution for the net benefit of a searched set Γ_{ik} , we need to add an additional step to the estimation approach. Therefore, the simultaneous search model under the assumption of normally distributed prices is estimated the following way: First, we take Q draws from ϵ_{ij} for each consumer/company combination. Second (new step), for each ϵ_{ij} draw, we take D draws from the price distributions for each consumer/company combination and calculate the expected maximum utility of a searched set as the average across all D draws.¹⁸ We repeat this step for each ϵ_{ij} draw. Third, for each ϵ_{ij} draw, we calculate the smoothed consideration and conditional purchase probabilities using a multivariate scaled logistic cumulative distribution function (CDF) (Gumbel 1961) with scaling parameters $s_1 = \dots = s_M = 5$. Fourth, we average the smoothed consideration and conditional purchase probabilities across all ϵ_{ij} draws. In the estimation, we set D to 200 and Q to 100.

4.2. Sequential Search

Since we do not observe the sequence of searches, we point out that observing a consumer's consideration set allows us to draw two conclusions based on Weitzman's (1979) rules: First, the minimum reservation utility among the searched companies has to be larger than the maximum reservation utility among the nonsearched companies (based on the selection rule), i.e.,

$$\min_{j \in S_i} r_{ij} \geq \max_{j' \notin S_i} r_{ij'}. \quad (14)$$

Otherwise, the consumer would have chosen to search a different set of companies. Second, the stopping and choice rules in Equations (5) and (6) can be combined to the following condition:

$$\max_{j \in S_i} u_{ij} \geq u_{ij'}, \quad \max_{j' \notin S_i} r_{ij'}, \quad \forall j' \in S_i \setminus \{j\}, \quad (15)$$

i.e., that the maximum utility among the searched companies is larger than any other utility among the considered companies *and* the maximum reservation utility among the nonconsidered companies.

Equations (14) and (15) are conditions that have to hold based on Weitzman's (1979) rules for optimal behavior under sequential search and *given the search and purchase outcome* that we observe in the data. At the same time, it must also have been optimal for the consumer not to stop searching and purchase earlier given

Weitzman’s (1979) rules. The challenge, as specified in the second issue raised at the beginning of this section, is that we do not observe the order in which the consumer collected the price quotes. The critical realization is that, *given the parameter estimates*, the observed behavior must have a high probability of having been optimal.

To illustrate, suppose a consumer searches three companies. Then the parameter estimates also have to satisfy the conditions under which it would have been optimal for the consumer to continue searching after his first and second search. Formally, in the estimation, given a set of estimates for the unknown parameters, for each consumer i , let us rank all searched companies j according to their reservation utilities \hat{r}_{it} (the “ $\hat{\hat{}}$ ” symbol refers to quantities computed at the current set of estimates) where $t = 1, \dots, k$ indicates the rank of a consumer’s reservation utility among the searched companies. Note that $t = 1$ ($t = k$) denotes the company with the largest (smallest) reservation utility \hat{r}_{it} . Furthermore, rank all utilities of searched companies in the same order as the reservation utilities; i.e., $\hat{u}_{i,t=1}$ denotes the *utility* for the company with the highest reservation utility $\hat{r}_{i,t=1}$. Then, *given the current parameter estimates*, the following conditions have to hold:

$$\hat{u}_{i,t=1} < \hat{r}_{i,t=2} \cap \max_{t=1,2} \hat{u}_{it} < \hat{r}_{i,t=3}.$$

In other words, although the *reservation utility* of the company with $t = 1$ is larger than that with $t = 2$ by definition, the *utility* of the company with $t = 1$ is smaller than the *reservation utility* of the company with $t = 2$, thereby prompting the consumer to do a second search. Similarly, the maximum utility from the (predicted) first and second search has to be smaller than the reservation utility from the (predicted) third search, otherwise the consumer would not have searched a third time. Generally, for a consumer searching $t = 2, \dots, k$ companies, the following set of conditions has to hold:

$$\bigcap_{l=2}^k \max_{t<l} \hat{u}_{it} < \hat{r}_{i,t=l}. \quad (16)$$

To calculate a consumer’s reservation utilities, we follow the approach suggested by Kim et al. (2010). The additional estimation conditions as described in Equation (16) are necessary to correctly recover search costs. These conditions impose restrictions on the utilities and bound the search cost parameter from above. Without these conditions, the search cost estimate is biased upward. We describe the reason for this bias in Section 5.4.2.

Since in the sequential search model, in contrast to the simultaneous search model, there are no separate consideration and conditional purchase stages, the probability of observing a consumer search a set

of companies Υ and purchase from company j under sequential search is

$$P_{ij\Upsilon|\epsilon} = P \left(\min_{j \in S_i} r_{ij} \geq \max_{j' \notin S_i} r_{ij'} \cap \max_{j \in S_i} u_{ij} \geq u_{ij''}, \max_{j' \notin S_i} r_{ij'} \cap \bigcap_{l=2}^k \max_{t<l} \hat{u}_{it} < \hat{r}_{i,t=l}, \forall j'' \in S_i \setminus \{j\}, t = 2, \dots, k \right). \quad (17)$$

Then, the log-likelihood of the model is given by

$$L = \prod_{i=1}^N \int_{-\infty}^{+\infty} \prod_{l=1}^L \prod_{j=1}^J P_{ij\Upsilon|\epsilon}^{y_{il}} f(\epsilon) d\epsilon, \quad (18)$$

where y_{il} indicates the chosen consideration set and the purchased company. In principle, we can write out all rankings of utilities and reservation utilities that satisfy the conditions in Equation (17) and write the probability of observing a consumer’s search and purchase behavior by calculating the sum of the probabilities of all admissible rankings. The challenge with writing out all utility and reservation utility rankings that satisfy the conditions in Equation (17) is that their number and complexity increases very quickly with the number of searches a consumer makes. Since, in our empirical application, we observe consumers searching up to 10 times, this approach is not feasible. A second challenge is that, even if we wrote out all admissible rankings of utilities and reservation utilities, the probability as described in Equation (17) does not have a closed-form solution. We use SMLE to estimate the sequential search model as it allows us to overcome both challenges. SMLE does not solve the combinatorial problem, but it circumvents it by allowing us to estimate the probability of observing a consumer search a set of companies Υ and purchase from company j in Equation (17) without having to write out all admissible rankings.

As in the estimation of the simultaneous search model, we use a kernel-smoothed frequency simulator (McFadden 1989) and smooth the probabilities using a multivariate scaled logistic CDF (Gumbel 1961). We describe the details of our estimation approach in Online Appendix A. As we describe in the next section, we have assessed the performance of our estimators only in a simulation context. Evaluating the theoretical properties of the estimators is left for future research; here we appeal to the literature on simulation estimators (see, e.g., Hajivassiliou 2000) to justify our choice of estimation strategy.

4.3. Monte Carlo Simulations

We conduct a large set of simulation studies to illustrate search method identification, demonstrate our estimation approaches, and evaluate consequences on

estimates, model fit, etc., of making incorrect assumptions for the search method, price distributions, and the search costs in Online Appendix C. Here, we briefly summarize our results from these simulation studies: First, the pattern of actual prices in consideration sets identifies the search method consumers use. Next, our estimation approaches are able to recover consumer preferences and search costs well when the true data generating process is known. Furthermore, when the wrong search method assumption is imposed in the estimation, true consumer preferences and search costs are no longer recovered. Especially the price and search cost coefficients are biased downward. We provide the intuition behind those downward biases in Online Appendix D. We also show that the model fit under the incorrect assumption on the search method is worse than the model fit under the correct assumption on the search method. We take this as evidence that the fit statistic is an additional suggestive predictor of the correct search method.¹⁹ Finally, making the correct assumption on the variance of the price distributions and that search costs are not company-specific is crucial to recover true consumer preferences and search costs—especially for the simultaneous search model.

5. Empirical Application

We use data on consumer search and purchase behavior for auto insurance from an insurance shopping study conducted by a large marketing research company in 2006 and 2007. We observe which companies consumers collected price quotes from and which companies consumers signed up with. This gives us information on consumers' consideration sets and purchase decisions. In addition, we observe monthly company-specific TV and radio advertising spending, consumer- and company-specific advertising recall, and quoted prices. We also have data on demographic variables, psychographic factors, and observed consumer attitudes toward insurance companies. As noted before, since consumers in our sample have been insured previously and coverage levels tend not to change much, assuming that consumers engage in price search and not in the search for other "attributes" is a reasonable assumption in this context.

Table 1. Descriptive Statistics

Variable	N	Mean	Std. dev.	Minimum	Maximum
Number of Price Quotes	945	2.96	1.38	1	10
Premium for 6-Month Policy with Current Insurer	945	592.97	288.28	74	2,750
Number of Vehicles	945	1.58	0.64	1	3
Number of Drivers	945	1.64	0.59	1	4
Vehicle Year	945	2001.98	4.19	1960	2007
Respondent Age	945	45.23	12.94	20	84

Note. Prices are measured in dollars.

5.1. Data Description

Table 1 contains descriptive statistics of our data, and Figure 2 shows a histogram of consumers' consideration set sizes. Consumers get, on average, 2.96 quotes (including one from their previous insurer), with the majority of consumers collecting two or three quotes (see Figure 2). The average premium with the current insurer is \$592.97 (see Table 1). Table 2 compares the mean characteristics for each respondent type (no search/no switch, search/no switch, search/switch). Consumers who neither search nor switch get, as expected, only one quote—the one that their previous insurer sends to them. Consumers who search, but decide not to switch collect 2.89 quotes on average, and consumers who search and switch gather 3.51 quotes. Consumers who neither search nor switch pay the highest average premium (\$660.13), followed by consumers who search but decide to stay with their previous insurer (\$606.36). Consumers who search and switch pay, on average, \$551.44. For further details, we refer the reader to Honka (2014) for a more detailed description of our data.

5.2. Empirical Model

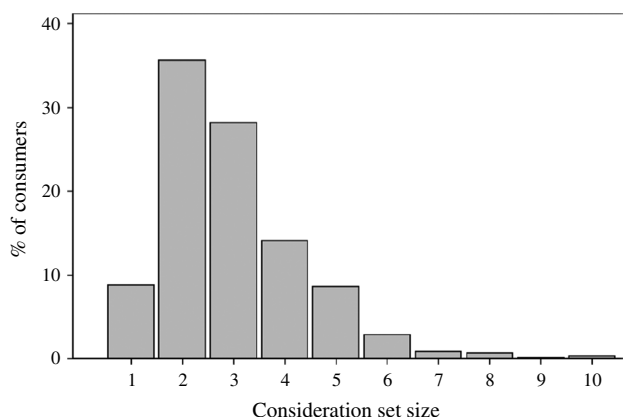
We assume consumers have rational expectations about prices and estimate consumers' price expectations using prices charged by previous insurers and a large set of variables that determine insurance prices such as demographics, drivers, cars, location, past claims history, other insurance products, and coverage choices.²⁰ We assume that prices follow a normal distribution with the mean being a function of the variables that determine insurance prices and a constant variance. The estimation results for the pricing regression are shown in Table 3. We use the predicted prices from this regression as price expectations in the main model estimation. Note that within a consumer, the expected prices across firms only vary because of the company-specific fixed effects.²¹ We refer the reader to Honka (2014) for details on the price expectation estimation process.

Consumer's utility for auto insurance is given by

$$u_{ij} = \alpha_j + \beta_1 p_{ij} + \beta_2 \text{adv}_{ij} + \beta_3 I_{ij,t-1} + Z_{ij} \gamma + \epsilon_{ij}, \quad (19)$$

where adv_{ij} denotes consumer- and company-specific recalled advertising. It is calculated as an interaction

Figure 2. Consideration Set Sizes



effect between consumer- and company-specific advertising recall and company-specific advertising spending. The term $I_{ij,t-1}$ is a dummy variable indicating whether consumer i made a purchase from the same company j as in time period $t - 1$, and Z_{ij} are observed psychographic factors. Collectively, $I_{ij,t-1}$ and Z_{ij} account for state-dependence and heterogeneity. With an average retention rate of about 70% in the auto insurance industry, capturing consumer inertia through β_3 is necessary to fully describe consumer behavior in this market. Furthermore, we also control for the following four psychographic factors denoted by Z_{ij} : *attitude toward auto insurance shopping and switching*, *new technology adoption*, *proven reliability*, and *out-of-box character*. While the first two variables are consumer-specific, the last two variables (*proven reliability* and *out-of-box character*) are both consumer- and company-specific.²² We chose to include these four factors because Honka (2014) has shown that they significantly influence consumers’ utility for auto insurance. Note that, under sequential search, the effects of consumer-specific variables in the utility function cannot be identified (see discussion in Section 5.4.2). We will therefore also explore the effects of these variables on search costs by making search costs a function of the psychographic factors.

Table 2. Averages Across Customer Types

Variable	No search/ No switch	Search/ No switch	Search/ Switch
Number of Respondents	56	586	303
Number of Price Quotes	1	2.89	3.51
Premium for 6-Month Policy with Current Insurer	660.13	606.36	551.44
Number of Vehicles	1.61	1.60	1.54
Number of Drivers	1.73	1.65	1.60
Vehicle Year	2002.93	2001.86	2002.05
Respondent Age	42.17	46.90	42.35

Note. Prices are measured in dollars.

It is common practice in the auto insurance industry that consumers receive a renewal offer about one month before their policy is set to expire. We view this renewal offer as a “free” first search since the consumer does not have to exert any effort to receive the price quote. Furthermore, we assume that a consumer knows the price his previous insurer is going to charge him to renew his insurance policy before making the decision (not) to search other companies. Finally, we assume the search set S_i contains all companies the consumer actively searches and the consumer’s consideration set C_i contains all searched companies and the previous insurer, i.e., $C_i = S_i \cup \{j_{I_{ij,t-1}}\}$.

5.3. Model-Free Evidence of Search Method

It is important to recognize that the (free) price quotes from previous insurance providers constitute a setting in the auto insurance industry that is nonstandard for consumer search models and specific to this industry. This is so because consumers do not “optimally” (under the rules of optimal consumer search) choose the company from which they get the first price quote, but it is always their previous insurance provider. If consumers were able to optimally choose which company to search first, they might or might not first search their previous insurance provider. This nonstandard setting, i.e., the fact that the first search might or might not be optimal, needs to be taken into account when identifying the search method consumers use when searching for prices of auto insurance policies. In general, the search method is not identified in such a nonstandard setting without an additional assumption.²³ Therefore, we make the additional assumption that consumers would have chosen to always include their previous provider in their consideration set (simultaneous search) or first search their previous insurance provider (if they could make that decision under sequential search). In the latter case, the price quote from their previous insurance provider is an optimal first search for consumers. The advantage of this particular assumption is that we can empirically *test* its appropriateness in our data. This is due to the particular pattern in our data and might not be possible with other data.

5.3.1. Pattern Under Simultaneous Search. Let us start out by discussing the pattern of actual prices in consumers’ consideration sets that is expected under simultaneous search when consumers first receive price quotes from their previous insurers and then decide optimally whether and which company(ies) to search in addition. The price quotes from the previous insurance providers influence consumers’ decisions to search beyond the previous insurer price quotes. Note that in the best case scenario, i.e., when consumers optimally choose which companies to search, the proportion of below-expectation actual prices among consumers searching once is λ . When the first price quotes

Table 3. Price Distribution

Variable	Estimate	Std. error	Variable	Estimate	Std. error
Constant	2.2886**	(0.7624)	21st Century	0.2374	(0.5118)
Male	-0.2063	(0.1574)	AIG	0.4593	(0.4107)
Married	-0.8474**	(0.2718)	Allstate	-0.1791	(0.3049)
Divorced/Separated	-0.1308	(0.2635)	American Family	-0.6157	(0.5088)
Widowed	0.3808	(0.7642)	Erie	-1.3207***	(0.5062)
Domestic Partnership	-0.1202	(0.3344)	Farmers	-0.3980	(0.3936)
Age	-0.0224***	(0.0065)	Geico	-1.0452***	(0.2972)
Driver under 25 Years	1.0810**	(0.3316)	GMAC	0.6378	(0.6153)
Two Vehicles	2.0683***	(0.1986)	Hartford	-0.7291	(0.4461)
Three Vehicles	4.1097***	(0.3031)	Liberty Mutual	0.3033	(0.3907)
Two Drivers	0.3489	(0.2683)	Mercury	0.2958	(0.5137)
Three Drivers	2.0738***	(0.4954)	MetLife	0.9811	(0.5057)
Four Drivers	1.3158	(0.9044)	Nationwide	-0.4867	(0.3926)
Medium City Suburb	0.0366	(0.2498)	Progressive	-0.1205	(0.3162)
Large City Suburb	0.6730**	(0.2491)	Safeco	0.2390	(0.5773)
Urban Area	1.0057***	(0.2776)	Travelers	0.7033	(0.4416)
Home Owner Insurance with Current Insurer	-0.1854	(0.1720)	Chosen Coverage	Yes	
Other Insurance with Current Insurer	-0.2136	(0.1746)	State	Yes	
Two or More Accidents	2.7329***	(0.4659)	Make × Class	Yes	
Two or More Tickets	1.2601***	(0.3656)			
Model Age	-0.0566**	(0.0182)	R ²	0.72	

Note. Prices are measured in \$100.

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

do not come from companies that are optimal to be searched for consumers, we expect the proportion of below-price-expectation actual prices among consumers who do not search beyond the previous insurer quote to be equal to or larger than the probability of getting a below-price-expectation price draw, λ . Intuitively speaking, this is the case because price quotes from “worse” than the optimal companies need to be lower than those from the optimal companies to prevent a consumer from searching further.²⁴

To summarize, under simultaneous search, if it were optimal for consumers to request price quotes from their previous insurance providers (if consumers can freely make that decision), the proportion of below-price-expectation actual prices among consumers who do not search beyond the previous insurers’ price quotes would equal the probability of getting a below-price-expectation price draw, λ . If it were not optimal for consumers to first request price quotes from their previous insurance providers (if consumers can freely make that decision), the proportion of below-price-expectation actual prices among consumers who do not search beyond the previous insurers’ price quotes would be larger than the probability of getting a below-price-expectation price draw, λ .

Conditional on the decision to search beyond the price quotes from the previous insurers, if consumers search simultaneously, the proportion of below-price-expectation actual prices in consumers’ search sets, i.e.,

consideration sets minus the price quote from the previous insurance provider, should equal the probability of getting a below-price-expectation price draw, λ .

5.3.2. Pattern Under Sequential Search. Under sequential search, even if the first price quotes do not come from companies consumers would have optimally chosen, consumers nevertheless react to the first price quotes under the rules of optimal sequential search, i.e., they stop searching when the price is “low enough” and continue searching when the price is “too high.”²⁵ The exact proportion of below-price-expectation actual prices among consumers who do not search beyond previous insurers’ price quotes will depend on the arrival process of the first price quotes, consumer preferences, etc. However, similar to the situation under simultaneous search, it is easy to recognize that the proportion of below-price-expectation actual prices among consumers who do not search beyond the previous insurers’ price quotes will be larger than or, at the limit, equal to the proportion of below-price-expectation actual prices under optimal sequential search, i.e., when consumers optimally pick the companies to be searched first, among consumers searching once.²⁶

5.3.3. Patterns in Our Data. Suppose a researcher has data from the auto insurance industry in which he observes that the proportion of below-price-expectation actual prices among consumers who do not search beyond the previous insurers’ price quotes is

Table 4. Proportion of Below-Expectation Prices

Percentage of below-expectation prices in search sets of size								
1	2	3	4	5	6	7	8	9
48.49 (2.74)	40.54 (3.15)	48.87 (4.60)	49.28 (6.02)	57.50 (10.09)	33.33 (17.82)	52.38 (20.39)	37.50 (48.41)	44.44 (49.69)

Note. Standard errors are in parentheses.

larger than the probability of getting a below-price-expectation price draw. Then the researcher cannot distinguish between (a) consumers searching simultaneously and getting a “nonoptimal” first price quote and (b) consumers searching sequentially (whether they get an “optimal” or “nonoptimal” first price quote). Thus, the search method is not identified.

Now suppose a researcher has data from the auto insurance industry in which he observes that the proportion of below-price-expectation actual prices among consumers who do not search beyond the previous insurers’ price quotes equals the probability of getting a below-price-expectation price draw. Such an empirical data pattern can only arise when consumers search simultaneously and getting a price quote from the previous insurance provider is “optimal.” In such a case, we have found support for our additional assumption that the price quotes from the previous insurance providers represent an optimal search for consumers and that consumers search simultaneously.

In our data, we observe the proportion of below-price-expectation actual prices among consumers who stop searching after the previous insurer quotes to be 0.48, and we cannot reject the null hypothesis that this proportion equals 0.5 at $p < 0.05$.²⁷ This result suggests that consumers would search their previous insurance providers and not different companies if they could freely choose. To put it differently, this result suggests that our additional assumption for the empirical context of the auto insurance industry that consumers would have searched their previous insurance providers if they could make that decision is supported by (model-free) patterns in the data.

Next, we calculate the proportions of below-price-expectation actual prices in consumers’ search sets, i.e., consideration sets less the price quotes from the previous insurance providers, and the results are shown in Table 4. Note that we have less than 10 observations for each of the search sets of sizes 6, 7, 8, and 9, and less than 30 observations for the search sets of size 5. Thus, we implement the chi-square test below twice: once using all search set sizes and once using search set sizes for which we have more than 30 observations, i.e., search sets of sizes 1 to 4.

Visually, the proportion of below-price-expectation actual prices in consumers’ search sets is around 0.5

(with some variation across search set sizes). We conducted *t*-tests (one for each search set size) and were not able to reject the null hypothesis that the proportion of below-price-expectation actual prices is 0.5 for all search set sizes with the exception of the search set of size 2 at $p < 0.05$. Then, additionally, we also tested the null hypothesis that the proportion of below-price-expectation actual prices in search sets of size 1 is the same as the proportion of below-price-expectation actual prices in search sets of size 2. We were not able to reject that null hypothesis at $p < 0.05$. Last, we first implemented a chi-square test that tested the null hypothesis that all proportions of below-price-expectation actual prices across the different search sets of sizes 1–9 are equal. Then, we also implemented a chi-square test that tested the null hypothesis that the proportions of below-price-expectation actual prices across search sets of sizes 1–4 are equal. We were not able to reject the null hypotheses that all proportions are equal, in both chi-square tests at $p < 0.05$. Thus we conclude that the pattern is consistent with consumers searching simultaneously.

To summarize, the search method consumers use can only be identified in an empirical context such as the one described for the auto insurance industry if we make the additional assumption that it would have been optimal for consumers to search their previous insurance providers (if consumers could make that decision). We find empirical support for this assumption in our data. Furthermore, we find that the patterns in our data show that consumers search simultaneously.

5.4. Model Parameter Identification

5.4.1. Under Simultaneous Search. We provide a brief summary of the discussion of identification of the model parameters under simultaneous search and refer the reader to Honka (2014) for more details. The identification of the parameters capturing differences in brand intercepts and other variables that vary across companies such as advertising spending is standard as in a conditional choice model. These parameters also play a role in consumers’ consideration set decisions.

The size of a consumer’s consideration set will help pin down search costs. We can only identify a range of search costs as it is utility-maximizing for all consumers with search costs in that range to search a

specific number of times. Beyond the fact that a consumer's search cost lies within a range that rationalizes searching a specific number of times, the variation in our data does not identify a point estimate for search costs. The search cost point estimate will be identified by the functional form of the utility function and the distributional assumption on the unobserved part of the utility.

Recall that we assume that the first search is free. The base brand intercept is identified from the consumer's decision to search or not to search beyond the free first search. Intuitively speaking, the free first search assumption creates a "fall-back option" similar to the outside option and allows us to identify the base brand intercept. So while the search cost estimate is pinned down by the average number of searches, the base brand intercept is identified by the search or no search decision (beyond the free first search). This also means that if there was a "fixed" component of the search cost that did not vary by the number of searches, this fixed cost would not be separately identified from the base brand intercept mean.

Consumer-specific variables that do not vary across companies are identified by consumers with certain characteristics searching more or less than others. For example, suppose older consumers search less than younger consumers. Then—given that the search cost coefficient is identified by the average number of searches across all consumers—older consumers must have a smaller benefit of searching, i.e., a lower utility for insurance, than younger consumers. Thus, we would expect a negative coefficient for age in the utility function. It is important to recognize that this argument only holds under the assumption of identical search costs across consumers. Alternatively, we could allow the consumer-specific variables to shift search costs instead of the utility.

5.4.2. Under Sequential Search. In the sequential search model, the parameters capturing differences in brand intercepts and variables that vary across companies such as advertising spending are identified from the conditions on the utilities and reservation utilities, i.e., Equations (4)–(6).

Search costs are identified from Weitzman's (1979) stopping rule (Equation (5)). They are not identified from the search rule, as it only imposes a *relative* ranking on the reservation utilities. Recall that the reservation utility is the utility that makes a consumer indifferent between searching and not searching. If there is a unique solution for Equation (4), as has been shown by previous research (e.g., Kim et al. 2010), and search costs are not company-specific as we assume in our empirical model, then the *relative* ranking of the reservation utilities will not change when search costs equally increase or decrease for all companies. Thus, search costs are not identified from Weitzman's (1979)

search rule. Search costs are also not identified from Weitzman's (1979) choice rule (Equation (6)), as search costs do not enter it. Search costs are identified by the stopping rule only as it describes the relationship between utilities and reservation utilities.

Previous research (e.g., Kim et al. 2010) has shown that reservation utilities decrease when search costs increase. Thus, as search costs increase, the stopping rule demanding that the maximum utility among the searched companies is larger than the maximum reservation utility among the nonsearched companies is satisfied earlier and consumers stop searching earlier. This is the mechanism behind the intuitive result that higher search costs make consumers search less. The number of searches a consumer makes identifies a range of search costs as it is utility-maximizing for a consumer with search costs in that range to search a specific number of times. For example, suppose it is optimal for a consumer to search once if his search costs lie between two and three, twice if his search costs lie between one and two, and three times if his search costs lie between zero and one. Then, by observing the consumer stop after the second search, we know that his search cost must be at least one, but we do not know whether his search costs are one, two, or three. Thus, imposing the stopping rule as shown in Equation (5) on the observed consideration set only puts a lower bound on the search cost estimate, as it only requires that search costs must have been larger than a lower bound to make the consumer stop searching. As a consequence, if only the stopping rule on the *observed* consideration set is used in the estimation, the search cost estimate exhibits an upward bias. This is the upward bias on the search cost estimate we described in Section 4.2.

By imposing the conditions that, given the current estimates, it must have been optimal for the consumer to continue searching (Equation (16)), we impose an upper bound on the search cost estimate that eliminates the previously described upward bias of the search cost estimate and allows us to recover the true values. The intuition here is that if the search costs had been higher, the consumer would not have continued searching. Beyond that a consumer's search cost lies within this range, which rationalizes stopping after a specific number of searches (but not earlier), the variation in our data does not identify a point estimate for search costs. The search cost point estimate will be identified by the functional form of the utility function and the distributional assumption on the unobserved part of the utility (as in the case of the simultaneous search model).

The base brand intercept—as in the simultaneous search model—is identified by a consumer's decision to search or not to search more than once given our assumption that the first search is free. Thus, observing a proportion of consumers to only search once (and

“pay” no search costs) is crucial in identifying the base brand intercept. As in the simultaneous search model, if there was a “fixed” component of the search cost that did not vary by the number of searches, this fixed cost would not be separately identified from the base brand intercept mean.

By contrast to the simultaneous search model, variables in the utility function that do not vary across companies, i.e., are consumer-specific, are not identified in the sequential search model. The effects of these consumer-specific variables are not identified from the choice or search rules as adding a constant to all utilities or reservation utilities does not change the relative rankings among the utilities or reservation utilities, respectively. The effects of these consumer-specific variables are also not identified from the stopping rule as adding a constant to the utility function does not affect the relationship between utilities and reservation utilities, i.e., whether a specific utility or a specific reservation utility is larger. The intuition behind this result is the following: Based on Kim et al. (2010), we know that a reservation utility in our model can be calculated as the sum of expected utility (expectation taken with respect to price) and a constant that depends on search costs, the price coefficient, and the standard deviation of the price distribution. Thus, for the same company j , any difference in utility for company j and reservation utility for company j comes from the difference in expected and actual price. For different companies, any difference in utility for company j and reservation utility for company j' comes from the difference in actual price for company j and expected price for company j' and differences in company-specific observed variables. Thus, variables that do not vary across companies do not affect the relationship between utilities and reservation utilities and are not identified from the stopping rule.

The lack of identification of the effects of variables that do not vary by alternative in the utility function in the sequential search method raises the issue of how to introduce demographic characteristics in models of search. For the simultaneous search model, these variables can be introduced either directly in the utility function or as shifters of search costs across consumers. For the sequential search model, only the latter operationalization is feasible. In the robustness checks (Section 7.1), we explore the consequences of introducing demographics—either in the utility function or in the search cost.

5.5. Estimation and Results

We need to adapt the estimation of both the simultaneous and the sequential search model compared to the ones shown in Section 4 to reflect a specific setting of the auto insurance industry, namely, that consumers know the prices their previous insurance providers

are going to charge them. For the simultaneous search model, we refer the reader to Honka (2014) for the estimation details. For the sequential search model, we refer the reader to Online Appendix E.

As described in Section 4, we need to make the assumption of first-order stochastic dominance among the price distributions to use Chade and Smith (2005) and Honka (2014) to estimate the simultaneous search model. We do so by assuming that the variances of the company-specific price distributions are identical. We tested the appropriateness of this assumption in two ways: First, we conducted a Bartlett test to test whether the company-specific variances of the price residuals from our pricing regression are equal. We were not able to reject the null hypothesis that all company-specific price residual variances are identical. Second, we estimated the normal distribution of prices by making the variance parameter a function of company fixed effects in addition to the mean parameter being a function of consumer demographics, insurance policy characteristics, and company fixed effects. None of the firm fixed effects on the variance parameter was significant. We conclude that the assumption of identical price variances is appropriate in our context.

Column (i) in Table 5 shows the results under the assumption that all consumers search simultaneously, and column (ii) shows the results under the assumption that all consumers search sequentially. The simultaneous search model fits the data better than the sequential search model. As discussed in Section 4.3, the model fit is an additional predictor of the search method consumers use. Furthermore, both the search cost and the price coefficient estimates under sequential search are very small. As outlined in Online Appendix D, when the true data generating process is simultaneous search but a sequential search model is estimated, we expect both the search cost and the price coefficient estimate to be severely biased downward. Together with the model-free evidence from the previous section that also indicated that consumers search simultaneously, we find consistent evidence of simultaneous search being the search method consumers use when shopping for auto insurance. Given this overwhelming evidence, we treat simultaneous search as the actual search method consumers use when shopping for auto insurance in the following sections. Our search cost estimate per search is \$42.09.²⁸ Note that the search cost estimate under simultaneous search is similar to the one (\$41.81) found by Honka (2014) using the same data and the same model (Model 0 in her paper). The small difference in search cost estimates can be explained by the different assumption on the price distributions noted previously. Dahlby and West (1986) find search costs to range from \$131 to \$570 (adjusted for inflation and converted from Canadian dollars) for auto insurance. We attribute these higher

Table 5. Results

	Search method			
	(i)		(ii)	
	Simultaneous search		Sequential search	
	Estimate	Std. error	Estimate	Std. error
Brand preferences				
<i>21st Century</i>	-1.7019***	(0.2047)	0.3814**	(0.1449)
<i>AIG</i>	-1.1634***	(0.2005)	0.8831***	(0.1223)
<i>Allstate</i>	-1.5772***	(0.2241)	0.8214***	(0.1612)
<i>American Family</i>	-1.5522***	(0.2204)	0.9392***	(0.1631)
<i>Erie</i>	-1.8607***	(0.2164)	0.7687***	(0.1584)
<i>Farmers</i>	-1.8135***	(0.2064)	0.6096***	(0.1384)
<i>Geico</i>	-2.1390***	(0.2762)	0.0801	(0.1649)
<i>GMAC</i>	-1.7801***	(0.3297)	0.7602***	(0.1455)
<i>Hartford</i>	-1.6454***	(0.1879)	0.5208***	(0.1238)
<i>Liberty Mutual</i>	-1.5790***	(0.2023)	0.1565	(0.1541)
<i>Mercury</i>	-1.8209***	(0.2427)	0.3002	(0.2401)
<i>MetLife</i>	-1.5956***	(0.2369)	0.4384**	(0.1467)
<i>Nationwide</i>	-1.9770***	(0.1951)	1.0862***	(0.1318)
<i>Progressive</i>	-1.4661***	(0.2173)	-0.0955	(0.0500)
<i>Safeco</i>	-2.1445***	(0.2422)	0.3894**	(0.1242)
<i>State Farm</i>	-1.5930***	(0.2172)	0.9673***	(0.1455)
<i>Travelers</i>	-1.5936***	(0.1952)	0.8484***	(0.1340)
Other parameters				
<i>Price</i>	-0.4479***	(0.0446)	-0.0683***	(0.0029)
<i>Recalled advertising</i>	0.1279**	(0.0491)	0.1303***	(0.0119)
<i>Inertia</i>	0.7172***	(0.0745)	0.6274***	(0.0361)
<i>Search cost</i>	0.1885***	(0.0445)	0.0005***	(0.0001)
Psychographic factors				
<i>Attitude toward auto insurance shopping and switching</i>	-0.4075**	(0.1419)		
<i>New technology adoption</i>	-0.1604	(0.1413)		
<i>Proven reliability</i>	0.3431***	(0.0556)	0.0100	(0.0226)
<i>Out-of-box character</i>	0.1436**	(0.0527)	0.0723**	(0.0263)
Log-likelihood	-3,079.12		-4,571.58	
Akaike information criterion	6,208.24		9,193.16	
Bayesian information criterion	6,346.85		9,331.77	

Note. Prices are measured in \$100.

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

search costs to the data used by Dahlby and West (1986) having been collected partially in rural Canada in the 1970s, where the density of agents was lower, and to the introductions of calling centers and the Internet, which significantly lowered consumer search costs. For Medigap insurance, Lin and Wildenbeest (2015) estimate median search cost per insurer to be \$30. We conclude that our search cost estimate is in line with search cost estimates found by previous literature in the same and related categories.

6. Effects of an Incorrect Assumption on the Search Method

Most previous empirical research on consumer search has made an assumption of the search method consumers use (e.g., Mehta et al. 2003, Kim et al. 2010, Seiler 2013). In this counterfactual, we investigate the effects of an incorrect search method assumption on

predicted quantities that are typically of interest to researchers such as elasticities, consideration set, and purchase market shares. To do so, we use the results from Table 5 and predict search cost elasticities, consideration set, and purchase market shares under the correct assumption of simultaneous search and the incorrect assumption of sequential search.

6.1. Search Cost Elasticities

We predict the percentage change in companies' consideration and purchase market shares due to a 10% decrease in search costs using simulation methods. Note that the search cost elasticity for purchase across all companies (in terms of number of purchased policies) is zero in the auto insurance market because consumers are required to have auto insurance. Thus, the total number of purchased auto insurance policies does not vary with search costs. The company-specific search cost elasticities for purchase can be both positive

and negative. Two effects determine search cost elasticities: First, some companies are hurt by a decrease in search costs because a consumer searches more companies due to the lower search costs, there is more competition within this consumer’s consideration set, and the company no longer gets chosen by the consumer. Second, some companies benefit from a decrease in search costs as a consumer decides to search more companies due to the decrease in search costs and the company newly gets searched and purchased. All companies encounter both effects when search costs decrease, and the net effect determines whether the company-specific search cost elasticity is positive or negative. Our search cost elasticity estimates need to be interpreted in light of this trade-off for each company.

Under the correct simultaneous search assumption, a 10% (50%) decrease in search costs results in a 4.03% (29.31%) increase in the average number of actively searched companies, i.e., excluding the free quote from the previous insurance provider, while under the incorrect assumption of sequential search, a 10% (50%) decrease in search costs results in a 1.89% (13.08%) increase in the average number of actively searched companies. Thus, the effects of a decrease in search costs are underestimated in terms of the number of actively searched companies under the incorrect assumption on the search method. Table 6 shows percentage changes in the consideration and purchase market shares due to a 10% decrease in search costs. Columns (i) and (iii) display the results under the correct assumption of simultaneous search, while columns (ii) and (iv) display the results under the incorrect assumption of sequential search. First, the predicted changes in both the consideration set and purchase shares under the incorrect assumption of sequential search are smaller than those under the correct assumption of simultaneous search. This is expected, as the number of active searches increases by only 1.89% under sequential search, which means that only very few companies are newly added to consumers’ consideration sets and thus their purchase decisions stay largely the same. Second, the predicted changes, especially in consideration set shares, can be different in terms of direction and magnitude under the correct and incorrect assumption on the search method. For example, while the consideration set shares for 21st Century, Progressive, and Travelers are rather similar, the consideration set shares for American Family, Hartford, and Liberty Mutual are quite different.

6.2. Consideration Set and Purchase Market Shares

Table 7 shows the predicted consideration set and purchase market shares under both assumptions on the search methods. Note that consideration set and purchase market shares can be over- or underpredicted

Table 6. Search Cost Elasticities Due to a 10% Decrease in Search Costs

Company	(i)		(ii)		(iii)		(iv)	
	Consideration set shares		Purchase market shares		Consideration set shares		Purchase market shares	
	Sim. search	Seq. search	Sim. search	Seq. search	Sim. search	Seq. search	Sim. search	Seq. search
21st Century	0.14 (0.08)	0.13 (0.03)	0.47 (0.11)	0.00 (0.02)	0.14 (0.08)	0.13 (0.03)	0.47 (0.11)	0.00 (0.02)
AIG	-0.12 (0.10)	0.02 (0.03)	0.10 (0.08)	0.00 (0.02)	-0.12 (0.10)	0.02 (0.03)	0.10 (0.08)	0.00 (0.02)
Allstate	-0.10 (0.03)	-0.04 (0.03)	0.31 (0.06)	0.00 (0.02)	-0.10 (0.03)	-0.04 (0.03)	0.31 (0.06)	0.00 (0.02)
American Family	-0.26 (0.23)	0.15 (0.05)	0.24 (0.16)	0.00 (0.03)	-0.26 (0.23)	0.15 (0.05)	0.24 (0.16)	0.00 (0.03)
Erie	-0.02 (0.10)	0.02 (0.05)	0.05 (0.10)	0.00 (0.04)	-0.02 (0.10)	0.02 (0.05)	0.05 (0.10)	0.00 (0.04)
Farmers	-0.13 (0.10)	-0.04 (0.03)	0.14 (0.06)	0.00 (0.03)	-0.13 (0.10)	-0.04 (0.03)	0.14 (0.06)	0.00 (0.03)
Geico	-0.25 (0.05)	-0.16 (0.04)	-0.03 (0.07)	0.00 (0.03)	-0.25 (0.05)	-0.16 (0.04)	-0.03 (0.07)	0.00 (0.03)
GMAC	0.15 (0.10)	0.23 (0.03)	-0.27 (0.06)	0.00 (0.02)	0.15 (0.10)	0.23 (0.03)	-0.27 (0.06)	0.00 (0.02)
Hartford	-0.09 (0.19)	0.10 (0.04)	0.10 (0.12)	0.00 (0.04)	-0.09 (0.19)	0.10 (0.04)	0.10 (0.12)	0.00 (0.04)
Liberty Mutual	0.41 (0.14)	0.07 (0.03)	0.32 (0.13)	0.00 (0.03)	0.41 (0.14)	0.07 (0.03)	0.32 (0.13)	0.00 (0.03)
Mercury	0.26 (0.12)	0.17 (0.04)	-0.17 (0.10)	0.00 (0.02)	0.26 (0.12)	0.17 (0.04)	-0.17 (0.10)	0.00 (0.02)
MetLife	0.19 (0.11)	0.12 (0.04)	-0.01 (0.11)	0.00 (0.03)	0.19 (0.11)	0.12 (0.04)	-0.01 (0.11)	0.00 (0.03)
Nationwide	0.07 (0.06)	0.00 (0.03)	0.02 (0.12)	0.00 (0.02)	0.07 (0.06)	0.00 (0.03)	0.02 (0.12)	0.00 (0.02)
Progressive	-0.07 (0.09)	-0.12 (0.04)	-0.25 (0.04)	0.00 (0.03)	-0.07 (0.09)	-0.12 (0.04)	-0.25 (0.04)	0.00 (0.03)
Safeco	0.42 (0.06)	0.14 (0.03)	-0.12 (0.12)	0.00 (0.02)	0.42 (0.06)	0.14 (0.03)	-0.12 (0.12)	0.00 (0.02)
State Farm	0.08 (0.13)	0.00 (0.04)	-0.48 (0.16)	0.00 (0.03)	0.08 (0.13)	0.00 (0.04)	-0.48 (0.16)	0.00 (0.03)
Travelers	-0.05 (0.09)	-0.08 (0.04)	-0.10 (0.08)	0.00 (0.03)	-0.05 (0.09)	-0.08 (0.04)	-0.10 (0.08)	0.00 (0.03)

Note. Bootstrapped standard errors are in parentheses.

under the incorrect assumption of sequential search. With the exception of 21st Century the direction of the over- or underprediction is directionally the same for consideration set and purchase market shares. Companies with the largest overpredictions of their purchase market shares (in percent) under sequential search are AIG, Allstate, and Progressive. Note that the purchase market shares of the four largest insurance companies (Allstate, Geico, Progressive, and State Farm) are being overpredicted under sequential search. Companies with the largest underpredictions of their purchase market shares (in percent) under sequential search are GMAC, Hartford, Mercury, and Safeco. The substantial differences in these shares across search methods for some of the brands in the data suggest that the choice of a search strategy in a given empirical application can have a large impact on predicted shares. This under-

Table 7. Consideration Set and Purchase Market Share Predictions

Company	(i)		(ii)		(iii)		(iv)	
	Consideration set composition				Purchase market shares			
	Sim. search		Seq. search		Sim. search		Seq. search	
21st Century	5.10 (0.63)	4.53 (0.32)	4.26 (0.60)	4.35 (0.32)				
AIG	7.89 (0.55)	7.99 (0.72)	6.91 (0.57)	8.07 (0.78)				
Allstate	9.03 (0.85)	9.62 (0.11)	8.12 (0.85)	9.78 (0.09)				
American Family	2.69 (0.39)	2.29 (0.14)	2.57 (0.40)	2.26 (0.14)				
Erie	3.15 (0.44)	3.13 (0.27)	3.13 (0.45)	3.12 (0.27)				
Farmers	6.31 (0.72)	6.98 (0.37)	6.09 (0.73)	6.99 (0.39)				
Geico	10.26 (0.65)	10.85 (0.65)	10.66 (0.68)	11.20 (0.66)				
GMAC	3.84 (0.55)	3.12 (0.96)	3.71 (0.56)	2.97 (1.00)				
Hartford	8.04 (1.53)	6.78 (0.55)	8.16 (1.61)	6.67 (0.57)				
Liberty Mutual	5.87 (0.84)	5.39 (0.47)	5.92 (0.90)	5.32 (0.48)				
Mercury	2.93 (0.51)	2.38 (0.26)	2.94 (0.57)	2.34 (0.25)				
MetLife	4.04 (0.60)	3.81 (1.77)	4.04 (0.67)	3.70 (1.81)				
Nationwide	5.62 (0.40)	5.16 (0.38)	5.83 (0.45)	5.00 (0.39)				
Progressive	9.97 (0.50)	12.32 (0.70)	11.04 (0.54)	12.77 (0.72)				
Safeco	3.18 (0.73)	2.73 (0.61)	3.34 (0.80)	2.58 (0.64)				
State Farm	7.55 (0.57)	8.58 (0.56)	8.45 (0.62)	8.69 (0.59)				
Travelers	4.54 (0.32)	4.34 (0.17)	4.81 (0.34)	4.19 (0.19)				

Notes. Consideration set and purchase market shares are shown in percentages. Bootstrapped standard errors are in parentheses.

scores the importance of an approach to identify the search method in such applications.

7. Robustness Checks

7.1. Alternative Model Specifications

We check the robustness of our empirical results by estimating four alternative model specifications. As discussed in Section 5.4.2, the effects of variables that do not change across companies such as two of the psychographics factors are not identified in the utility function under sequential search. We therefore estimate a simultaneous and sequential search model where those psychographics factors enter through consumers' search costs instead of consumers' utility function. While not reported here, we find that

for the simultaneous search model the estimates and the log-likelihood are very similar to column (i) in Table 5. For the sequential search model, we find that, while the inclusion of the two psychographics factors increases the log-likelihood (as to be expected), the log-likelihood of the model under simultaneous search remains larger. Other parameter estimates for the sequential search model remain very similar to column (ii) in Table 5.²⁹

The sequential search model is also more flexible in that the assumption of first-order stochastic dominance among the price distributions is not necessary for estimation. Thus, we estimate the sequential search model using company-specific price variances. While not reported here, we find neither the parameter estimates nor the log-likelihood to change much com-

pared to column (ii) in Table 5. This is likely due to the company-specific price variances being similar in the empirical data. Finally, we also estimate a sequential search model with company-specific price variances and consumer-specific psychographic factors entering the model through search costs. While we find the log-likelihood to increase, this increase is due to the inclusion of additional psychographic factors. Furthermore, we find that the log-likelihood of this most flexible sequential search model is still much smaller than the one of the less flexible simultaneous search model (column (i) in Table 5). We conclude that our results are robust to alternative model specifications and that simultaneous search is the appropriate modeling assumption to describe consumer shopping behavior in the auto insurance industry.

7.2. Unobserved Heterogeneity

In the empirical application, heterogeneity across consumers is captured via observable characteristics such as psychographic factors as well as lagged choice (a la Guadagni and Little 1983). One concern could be the presence of unobserved heterogeneity in preferences, search costs, and search method. Given the nature of our survey data, accounting for unobserved heterogeneity would typically not be possible. However, we have access to variables that could affect the search method such as credit scores, but that are unlikely to directly affect the utility of the alternatives.³⁰ With information on these variables, we can appeal to a discrete-heterogeneity concomitant variable approach (see, e.g., Dayton and Macready 1988) to distinguish those who search sequentially (consumers with high credit scores) from those who search simultaneously (consumers with low credit scores). We estimate such a two-segment concomitant variable latent class model (Kamakura and Russell 1989) where one class of consumers searches simultaneously and the other class of consumers searches sequentially. While not reported here, we find that the size of the simultaneously searching consumer segment is 0.88, and that the model estimates are very similar to the ones from model (i) in Table 5. Search costs are estimated to be \$42.11. We therefore conclude that our results are robust to this form of unobserved heterogeneity in the data.

8. Limitations and Future Research

There are several limitations to our research. First, we assume that consumers have rational expectations about prices. A model that has information on consumer price expectations or is able to recover them would enable researchers to test the hypothesis of rational price expectations and compare it with other price expectation formation theories. A related issue in our empirical context is that we have access only to cross-sectional data from 945 survey respondents. To recover

the price expectations for these consumers, we need to make a specific functional form assumption on the relationship between prices and consumer characteristics. Having access to more data would free us from having to make this assumption. Second, our search method identification strategy for the case where there is unobserved heterogeneity in search costs across consumers and companies relies on the nonverifiable assumption that all consumers have positive search costs. It will be a fruitful avenue for future research to come up with a search method identification strategy that does not make this assumption. Third, our model implicitly assumes that consumers make one and only one decision about the search method they want to use (and the number of quotes they are going to collect under simultaneous search) before starting any search activity. In reality, consumers might go through multiple search stages. For example, a consumer might initially decide to collect two price quotes searching simultaneously and then, after learning about the two prices, decide to search sequentially, stop after three price quotes, and make a purchase. Developing such a multistage search model is left for future research. To carry out such analyses, however, researchers need to be equipped with more detailed data than those used in this paper. At the same time, the data we use are increasingly becoming available; the approaches proposed in this paper therefore allow us to make progress on answering important questions regarding the magnitudes of search costs and consequences of assumptions made in estimating models of search.

Fourth, our goal in this paper is to present search models that can be used in markets with any number of companies. To estimate the simultaneous search model in markets with a lot of alternatives, we have to assume that search costs are not company-specific. Note that this limitation has no implications for our search method identification strategy and that it can be overcome in markets with few alternatives by using a choice model approach a la Mehta et al. (2003). Finally, following the standard search literature, our model assumes that consumers search to resolve uncertainty about a single product characteristic, i.e., price. Yet in many contexts, consumers might search to learn about two or more product characteristics. For example, consumers might search to learn about coverage options and prices in the auto insurance industry. We leave it for future research to develop a model that allows consumers to search for multiple product characteristics.

9. Conclusion

In this paper, we explore whether the search method consumers use can be deduced in data where consumer purchases and consideration sets, but not the sequence of searches, are observed. We show analytically that the search method is nonparametrically

identified in those kind of data. Under simultaneous search, the average proportion of below-expectation price draws is constant across all consideration set sizes and equals the probability of getting a below-price-expectation price draw, while under sequential search, the proportion of below-expectation price draws among consumers searching once is larger than the probability of getting a below-price-expectation price draw.

We suggest a new estimation approach for the sequential search model where the researcher has access to individual-level data on consideration sets, purchases, and other characteristics, but not the sequence of searches. Our SMLE approach is able to overcome the challenge of the researcher not knowing the sequence of searches.

We apply our model and estimation approach to data from the U.S. auto insurance industry and find consumers to search simultaneously with search costs of about \$42. Using our estimates, we predict consideration set and purchase market shares under both search methods. We find consideration and purchase market shares of the largest insurance companies to be overpredicted under the incorrect assumption of sequential search.

Acknowledgments

The authors thank Bart Bronnenberg; Jean-Pierre Dubé; Ken Hendricks; Ali Hortaçsu; Jean-François Houde; Stephan Seiler; Maria Ana Vitorino; Matthijs Wildenbeest; the participants of the 2012 Marketing Science Conference, 2012 INFORMS Annual Meeting, 11th International Industrial Organization Conference, 9th Invitational Choice Symposium, 2013 Quantitative Marketing and Economics Conference, 2014 Econometric Society Meeting, and 2015 NBER Industrial Organization Winter Meeting; and seminar participants at the Yale School of Management, University of Minnesota, University of Maryland, University of Miami, Harvard Business School, Temple University, Texas A&M University, University of California San Diego, Dartmouth College, University of Kansas, Indian School of Business, UCLA, Chinese University of Hong Kong, University of Michigan, and Southern Methodist University for their comments. The authors are grateful to an anonymous company for providing them with the data. The feedback from two reviewers, an associate editor, and the senior editor is also gratefully acknowledged. The usual disclaimer applies.

Endnotes

¹Note that we use the terms “consideration set” and “search set” interchangeably in Sections 1–4 of this paper. Starting with our empirical application in Section 5, we introduce a distinction between both terms.

²A search cost is an information “cost” borne by a consumer to acquire information about a firm—usually in the form of time and effort required to obtain such information. It does not have to be a monetary cost.

³We discuss these assumptions and the extent to which our search method identification approach relies on them in Section 3.2.1.

⁴While we do not model the supply side in this paper, our demand model is consistent with several general equilibrium models that result in price uncertainty and consumer search, e.g., a pure strategy equilibrium where consumers do not know company costs, which translates into uncertainty about prices (Benabou 1993), or a mixed strategy equilibrium model where companies randomize prices (Burdett and Judd 1983).

⁵We present the most general model with consumer- and company-specific price distribution means. This includes simpler settings where the price distribution means are company-specific, consumer-specific, or marketwide.

⁶For expositional purposes, we show β and γ without any subscripts. They can be added when appropriate.

⁷Note that our search method identification proof does not rely on the existence of both types of consumers, as we discuss in detail in Online Appendix B.

⁸Exceptions are, for example, Rothschild (1974), De los Santos et al. (2015), and Matsumoto and Spence (2014).

⁹Details about the characteristic patterns under simultaneous and sequential search when consumers’ price expectations are *not* rational are available from the authors on request.

¹⁰Based on our discussion in Section 3.2, the researcher can test the null hypothesis of simultaneous search either through K t -tests that the proportion of below-(true) price-expectation price draws in consumers’ consideration sets equals λ and/or through a chi-squared test for the proportion of below-(true) price-expectation price draws being constant across consideration set sizes.

¹¹Note that for assumption (ii), i.e., that the first search is free, assumption (f), i.e., that consumers have nonzero search costs, still applies. In that case, it applies to all searches beyond the free first search.

¹²Details are available from the authors on request.

¹³We discuss in detail and provide an example of why the assumption of nonzero search costs is necessary in this case in Case 5 in Online Appendix C.

¹⁴Note that we do not require the search cost distribution to be continuous over its full range. We only require it to be continuous over the interval 0 to $A > 0$. Our search method identification goes through when a search cost distribution has support, e.g., from 0 to A and from B to C with $C \geq B > A > 0$.

¹⁵We study the effects of the equal price variance and identical search costs assumptions for both simultaneous and sequential search models in simulation studies in Online Appendix D.

¹⁶Note that in markets with few alternatives, i.e., markets where the curse of dimensionality is not severe, one could use a choice-model approach to estimate the simultaneous search model (for details, see Mehta et al. 2003, Honka 2014). In that case, the assumptions of first-order stochastic dominance among the price distributions and that search costs are not company-specific are not necessary.

¹⁷We chose the assumption of normally distributed prices instead of EV type I distributed prices for both simultaneous and sequential search models since it allows us to use the approach suggested by Kim et al. (2010) to calculate the reservation utilities under sequential search. The Kim et al. (2010) estimation approach for the reservation utilities cannot be used when prices follow an EV type I distribution.

¹⁸Note that we hold the set of D draws from the price distributions constant within an estimation. We also hold it constant across all 50 replications in the Monte Carlo simulations.

¹⁹The fit statistic is only a suggestive predictor of the correct search method since simulations cannot cover the entire space of variables, parameters, and all possible functional forms of utilities.

²⁰Note that Honka (2014) conducted extensive checks to ensure that using prices charged by previous insurers is a valid approach

to estimating the marketwide distributions of insurance prices and does not suffer from a selection bias.

²¹Note that because of the size of our data, we are not able to include company-specific effects of demographic variables in consumers' price expectations.

²²Details on these psychographic factors and summary statistics of the items that constitute them are available from the authors on request.

²³The technical reason for the necessity of an additional assumption is as follows: If consumers do not optimally choose the company to search first, we cannot derive a general mathematical representation of the proportion of below-price-expectation actual prices among consumers who do not search beyond the first quote without additional information/data or other additional assumptions—neither under simultaneous nor sequential search (we provide details why this is the case in Sections 5.3.1 and 5.3.2).

²⁴This applies to differentiated goods. We view and model auto insurance as a differentiated good.

²⁵Consumers stop searching and purchase when the maximum reservation utility among the searched companies is larger than the maximum reservation utility among the nonsearched companies (Weitzman 1979).

²⁶Note that the necessary condition that we observe consumers to search more than once is satisfied in our data.

²⁷Recall that insurance prices depend on consumer and policy characteristics. In estimating the distribution of prices, we account for this by making the expected price a function of these consumer and policy characteristics. Thus, when calculating the proportion of below-expectation prices, we compare actual prices in consumers' consideration sets to consumer- and company-specific expected prices.

²⁸We calculate search costs in dollars by dividing the search cost coefficient c by the price coefficient β_1 .

²⁹We also estimated a sequential search model in which search costs are a function of company-specific advertising spending in the month prior to the consumer's insurance purchase. While advertising spending significantly decreases search costs, the log-likelihood of such a sequential search model remains worse than the log-likelihood of the simultaneous search model reported in column (i) in Table 5. Note that we cannot estimate such a model under simultaneous search, as search costs cannot be company-specific in our estimation approach (Chade and Smith 2005).

³⁰Morgan and Manning (1985) find that using simultaneous search is more efficient for consumers when prices need to be gathered quickly. Chade and Smith (2006) and Kircher (2009) find that using simultaneous search is more efficient for consumers when the other side of the market might reject the consumer. Variables that translate into these two factors in the auto insurance market are potentially the timing of the price search process (close to the policy expiration date or weeks in advance), tickets and accidents in the past, and low credit scores.

References

Andrews R, Srinivasan T (1995) Studying consideration effects in empirical choice models using scanner panel data. *J. Marketing Res.* 32(1):30–41.

Baye M, Morgan J, Scholten P (2006) Information, search, and price dispersion. Hendershott T, ed. *Handbook on Economics and Information Systems* (Elsevier, Amsterdam), 323–376.

Benabou R (1993) Search market equilibrium, bilateral heterogeneity, and repeat purchases. *J. Econom. Theory* 60(1):140–158.

Bronnenberg B, Vanhonacker W (1996) Limited choice sets, local price response, and implied measures of price competition. *J. Marketing Res.* 33(2):163–173.

Burdett K, Judd K (1983) Equilibrium price dispersion. *Econometrica* 51(4):955–970.

Chade H, Smith L (2005) Simultaneous search. Working paper, Yale University, New Haven, CT.

Chade H, Smith L (2006) Simultaneous search. *Econometrica* 74(5):1293–1307.

Chen X, Hong H, Shum M (2007) Nonparametric likelihood ratio model selection tests between parametric likelihood and moment condition models. *J. Econometrics* 141(1):109–140.

Chen Y, Yao S (2016) Sequential search with refinement: Model and application with click-stream data. *Management Sci.* Forthcoming.

Chiang J, Chib S, Narasimhan C (1999) Markov chain Monte Carlo and models of consideration set and parameter heterogeneity. *J. Econometrics* 89(1–2):223–248.

Dahlby B, West D (1986) Price dispersion in an automobile insurance market. *J. Political Econ.* 94(2):418–438.

Dayton C, Macready G (1988) Concomitant-variable latent-class models. *J. Amer. Statist. Assoc.* 83(401):173–178.

De los Santos B, Hortaçsu A, Wildenbeest M (2012) Testing models of consumer search using data on Web browsing and purchasing behavior. *Amer. Econom. Rev.* 102(1):2455–2480.

De los Santos B, Hortaçsu A, Wildenbeest M (2015) Search with learning. *J. Bus. Econom. Statist.* Forthcoming.

DeSarbo W, Jedidi K (1995) The spatial representation of heterogeneous consideration sets. *Marketing Sci.* 14(3):326–342.

DeSarbo W, Lehman D, Carpenter G, Sinha I (1996) A stochastic multidimensional unfolding approach for representing phased decision outcomes. *Psychometrika* 61(3):485–508.

Gensch D, Soofi E (1995) Information-theoretic estimation of individual consideration set. *Internat. J. Res. Marketing* 12(1):25–38.

Guadagni P, Little J (1983) A logit model of brand choice. *Marketing Sci.* 2(2):203–238.

Gumbel EJ (1961) Bivariate logistic distributions. *J. Amer. Statist. Assoc.* 56(294):335–349.

Hajivassiliou V (2000) Some practical issues in maximum simulated likelihood. Mariano R, Schuermann T, Weeks M, eds. *Simulation-Based Inference in Econometrics: Methods and Applications* (Cambridge University Press, Cambridge, UK), 71–99.

Hauser J, Wernerfelt B (1990) An evaluation cost model of consideration sets. *J. Consumer Res.* 16(4):393–408.

Hong H, Shum M (2006) Using price distributions to estimate search costs. *RAND J. Econom.* 37(2):257–275.

Honka E (2014) Quantifying search and switching costs in the U.S. auto insurance industry. *RAND J. Econom.* 45(4):847–884.

Honka E, Hortaçsu A, Vitorino M (2016) Advertising, consumer awareness, and choice: Evidence from the U.S. banking industry. Working paper, University of California Los Angeles.

Janssen M, Moraga-Gonzalez J, Wildenbeest M (2005) Truly costly sequential search and oligopolistic pricing. *Internat. J. Indust. Organ.* 23(5):451–466.

Kamakura W, Russell G (1989) A probabilistic choice model for market segmentation and elasticity structure. *J. Marketing Res.* 26(4):379–390.

Kim J, Albuquerque P, Bronnenberg B (2010) Online demand under limited consumer search. *Marketing Sci.* 29(6):1001–1023.

Kircher P (2009) Efficiency of simultaneous search. *J. Political Econ.* 117(5):861–913.

Lin H, Wildenbeest M (2015) Search and prices in the Medigap insurance market. Working paper, Indiana University, Bloomington.

Matsumoto B, Spence F (2014) Price beliefs and experience: Do consumers' beliefs converge to empirical distributions with repeat purchases? Working paper, University of North Carolina at Chapel Hill, Chapel Hill.

McFadden D (1986) The choice theory approach to market research. *Marketing Sci.* 5(4):275–297.

- McFadden D (1989) A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica* 57(5):995–1026.
- Mehta N, Rajiv S, Srinivasan K (2003) Price uncertainty and consumer search: A structural model of consideration set formation. *Marketing Sci.* 22(1):58–84.
- Miller G (1956) The magic number seven, plus or minus two: Some limits on our capacity for processing information. *Psych. Rev.* 63(2):81–97.
- Mitra A, Lynch J (1996) Advertising effects on consumer welfare: Prices paid and liking for brands selected. *Marketing Lett.* 7(1):19–29.
- Moraga-Gonzalez J, Wildenbeest M (2008) Maximum likelihood estimation of search costs. *Eur. Econom. Rev.* 52(5):820–848.
- Morgan P, Manning R (1985) Optimal search. *Econometrica* 53(4):923–944.
- Muir D, Seim K, Vitorino M (2013) Price obfuscation and consumer search: An empirical analysis. Working paper, University of Pennsylvania, Philadelphia.
- Pires T (2015) Consideration sets in storable goods markets. Working paper, University of North Carolina at Chapel Hill, Chapel Hill.
- Ratchford B (1980) The value of information for selected appliances. *J. Marketing Res.* 17(1):14–25.
- Reinganum J (1979) A simple model of equilibrium price dispersion. *J. Political Econom.* 87(4):851–858.
- Roberts J, Lattin J (1991) Development and testing of a model of consideration set composition. *J. Marketing Res.* 28(4):429–440.
- Roberts J, Lattin J (1997) Consideration: Review of research and prospects for future insights. *J. Marketing Res.* 34(3):406–410.
- Rothschild M (1974) Searching for the lowest price when the distribution of prices is unknown. *J. Political Econom.* 82(4):689–711.
- Seiler S (2013) The impact of search costs on consumer behavior: A dynamic approach. *Quant. Marketing Econom.* 11(2):155–203.
- Shugan S (1980) The cost of thinking. *J. Consumer Res.* 7(2):99–111.
- Siddarth S, Bucklin R, Morrison D (1995) Making the cut: Modeling and analyzing choice set restriction in scanner panel data. *J. Marketing Res.* 32(3):255–266.
- Stahl D (1989) Oligopolistic pricing with sequential consumer search. *Amer. Econom. Rev.* 79(2):700–712.
- Stigler G (1961) The economics of information. *J. Political Econom.* 69(3):213–225.
- Stiglitz J (1987) Competition and the number of firms in a market: Are duopolies more competitive than atomistic markets? *J. Political Econom.* 95(5):1041–1061.
- Weitzman M (1979) Optimal search for the best alternative. *Econometrica* 47(3):641–654.