Guiding Consumers through Lemons and Peaches: 
A Dynamic Model of Search over Multiple 
Characteristics

Pedro M. Gardete     Megan H. Antill

May 2019

Abstract

The increasing amount of data available to consumers has most likely aided in decision-making. However, it has also created an opportunity for sellers to design the information landscape that consumers navigate. This paper develops a novel fully dynamic search model for alternatives with multiple characteristics, and reports estimation results for an online used car seller. The model allows characterizing search over alternatives with multiple characteristics that may be distributed arbitrarily. It also allows for a rich set of consumer search behaviors, including piecemeal search within and arbitrary paths across alternatives. We estimate the model using clickstream data on the website of a used car seller. The dataset tracks incremental search actions as well as test-drive reservations. The estimated fundamentals are then used to consider the effects of different information design policies. We find that the choice of the characteristics to be made available to consumers upfront may have conversion implications ranging from -0.39% to +1.65%. The perfect information scenario increases conversion rates by 10.4%. Finally, we compare our model with the knowledge gradient model of learning, and show that taking forward-looking behavior into account explains the moments of the data better, and that the models’ likelihoods are significantly different.

*Graduate School of Business, Stanford University, Stanford, California 94305. We thank Carlos Daniel Santos, Ilya Morozov, and Robert Sanders for helpful comments as well as the participants of research seminars at Catolica Lisbon, INSEAD, Nova School of Business and Economics, Stanford W.I.P., and UC San Diego. Correspondence: Gardete: gardete@stanford.edu, Antill: mkhunter@stanford.edu.
1 Introduction

In recent years, the increased availability of data surrounding the benefits and characteristics of sellers’ offerings has allowed consumers to make better-informed decisions. Consumers access diagnostic information across a variety of domains, from simple nutrition facts of candy bars to complex neighborhood characteristics of a potential new home. On their end, sellers have made a significant amount of data available to consumers: Real estate websites provide photo galleries and open house scheduling information and car dealership websites provide extensive information about the vehicles available in their lots.

The result of this demand and supply interaction for decision-aiding information is a large data ecosystem that consumers seek to browse efficiently. The recent literature on consumer search has made rapid progress in understanding how consumers navigate the information landscapes offered up by sellers. However, little is known about the implications of the information design decisions made by the sellers themselves on consumer search and ultimately, profitability.

In this paper, we investigate the implications of the design of information provision by a seller who faces consumers who search over products with multiple characteristics. We believe this question constitutes a focal counterfactual in the search literature. This paper models consumer search in order to recover the consumer-side fundamentals that then allow us to consider different information disclosure policies. Each of the counterfactual analyses takes into account that, when faced with different information disclosure policies, consumers reoptimize their own information procurement strategies. Our work relates to the notion of ‘search attributes’ as proposed by Nelson (1970). However, we allow the seller to endogenize which attributes should be known ‘ex ante’ to consumers and which should be made ‘search attributes.’

This paper offers three contributions to the literature on information search. First, we develop a model that allows for a rich characterization of forward-looking behavior of consumers searching over multiple characteristics. A distinctive feature of our unique dataset is that we observe incremental consumer search behaviors within each alternative. We observe multiple alternatives being searched only partially, before the consumer’s focus moves on to a different prospect. While this ‘piecemeal’ search behavior allows us to identify the implications of redesigning the information environment at the characteristic level, it also
demands a new modeling approach, since the observed behavior rules out classical models such as the Weitzman (1979) sequential search model.

Our model introduces a number of novel features to the search literature. First, it allows for piecemeal search of alternatives, as explained above, as well as flexible search behaviors, like returning to a previously inspected alternative in order to conduct additional search. Second, the model allows consumers to take into account that product characteristics are correlated, which means that learning about a given characteristic may affect posterior beliefs over the remaining characteristics, as well as subsequent search decisions. For example, in our setting (an online used car seller), consumers are allowed to take into account that older cars may have more issues and higher mileage.

We approximate the empirical joint distribution of product characteristics through the use of a statistical copula. This approach allows for a flexible estimation of the marginal distributions, as well as the characterization of the relationships among them. The recovered joint distribution of characteristics is used to inform consumer beliefs during estimation as well as during the counterfactual analyses. A third distinctive feature of our model is that it allows for an unobserved utility component, whose correlations with the observed characteristics can also be estimated. As we explain later, identification stems from state-dependent search patterns, as they vary as a function of the levels of the observable characteristics of the alternatives.

These innovations in modeling search are enabled by two key technical elements. First, we present a ‘state-bootstrapping’ estimator, which relies on drawing simulations from the joint distribution of vehicle characteristics in order to form the consumers’ value functions. This approach yields a trivially consistent simulated maximum likelihood estimator, and is especially useful for dynamic models of search. Second, we represent the value function of each consumer efficiently through a ‘state-partitioning decision tree’, where each search action taken by the decision-maker leads her to exclude possible cases in the state space. The combination of the two techniques is especially productive in search settings, where consumers’ continuation actions transition them to ‘smaller’ state spaces, i.e., the posterior distributions become tighter as more information is collected.

The second main contribution of this paper is to use the estimated fundamentals to inform the information provision strategy of the firm. Despite the rapid progress in characterizing consumer search, there is little knowledge on how decisions related to the information envi-
ronment affect firm profitability. An exception is the work by Ellison and Wolitzky (2012), who show that by obfuscating attributes, sellers can impose heavy burdens on consumers, who anticipate facing long higher future search costs before being able to retrieve relevant information. Given the large data ecosystem available to sellers, a potentially key decision is which information to make immediately available to consumers, and which information to relegate to be discovered through search actions. Consider, for example, the decision of moving information about a product characteristic from the main listing page down the page hierarchy. The reason the model described above is useful to analyze different information scenarios is that it takes into account that changes to the information disclosure policy affect consumers’ subsequent search decisions. After recovering the fundamental search and preference parameters, the model is used to consider counterfactual consumer search outcomes across different information disclosure scenarios.

It may appear counterintuitive that a redesign of the information layout can have supply-side implications, in light of the seller facing forward-looking consumers who hold correct beliefs about the distribution of characteristics. We find that such redesigns may indeed carry outcome implications. The intuition is as follows: If all consumers expect negative ex ante utility from the alternatives offered, then the seller has an incentive to provide some information. This action is likely to ‘partition’ demand: Some consumers will update positively based on the information they find (e.g., they may find an alternative that they like) whereas others may update negatively. As we illustrate later, we also find that the seller may become better off making consumers relatively indifferent between searching further or not. If at least some of these consumers have already found their matches, then the seller can limit subsequent ‘negative surprises’ by discouraging additional consumer search. In the context of high ticket items such as ours, the profitability effects of different information designs are modest, yet economically meaningful and statistically significant.\footnote{In our data, we do observe conversions taking place despite additional search actions being available to consumers.}

The third contribution of our paper is to compare the findings of our forward-looking model with the knowledge gradient literature, as proposed in the Operations and Economics literatures (see for example, Frazier, Powell, and Dayanik (2009), Powell (2010), and Liang, Mu, and Syrgkanis (2017)). The knowledge gradient literature considers myopic decision-makers who behave as if each search decision will lead them towards a terminal decision in
the next period. In this way, the decision-maker looks for the steepest ascent in option value induced by search at each moment, net of the associated search cost. The knowledge gradient model is extremely convenient for researchers, since it simplifies calculating the future value of search and the optimal search policies.

We modify our model to implement and estimate the knowledge gradient policy. We reject this policy as a good approximation to the forward-looking problem: The models exhibit statistically different likelihood values, and the knowledge gradient method fits the moments in the data less accurately than our model. The fact that dynamic learning models fit the moments in the data better is not always guaranteed (see for example Erdem and Keane (1996)), so our results highlight the need to at least consider forward-looking behavior when modeling consumer search.

One important aspect of modeling search is defining the set of alternatives consumers may consider evaluating. More parsimonious search models, like the Weitzman sequential model, allow for a large number of candidate alternatives (which we will refer to as consideration sets). However, a known limitation in search models is the fact that each consumer typically contributes with a single observation (i.e., a sequence of search decisions followed by a terminal action), which limits the degree to which heterogeneity is identified. As a result, in typical Weitzman search model applications, consumers search alternatives in the same order, but may differ in the realized consideration sets only because of random draws, learned during search. The underlying mechanism is intuitive: with homogeneous preference parameters, variation in consideration sets is explained only through random shocks.

 Appropriately modeling consideration sets is a frontier problem in the search literature, and we do not solve it in this paper. Rather than assuming that all consumers consider all of the alternatives and inspect them in the same order, we define a consumer’s consideration set as the set of products we observe her interact with in the data. This way, we introduce consideration set variation in the estimation, while keeping the number of search alternatives manageable. Our final working sample is obtained by focusing on the three top makes sold by the firm, which account for more than a third of the sedan vehicle stock. Moreover, we take advantage of the fact that 89% of consumers in our dataset evaluate at most 4 alternatives, which already generates \((2^4)^4 = 4,096\) possible search action configurations, and a state space with a much larger number of points (we discuss this later in more detail).\(^2\)

---

\(^2\)Modeling endogenous consideration sets would require computing an extremely demanding problem,
One innovation in our model is that, because we observe incremental search actions in our dataset, we are able to identify action-specific search costs. Moreover, we observed a great deal of improvement in fit after introducing a second latent segment of consumers, being faced with different search costs. Given the number of search actions we observe, in relation to the number of terminal decisions, we believe that introducing heterogeneity in the search parameters is relatively less demanding than introducing heterogeneity in preference parameters. Although identification is informed by functional form in this case, both AIC and BIC criteria favor the two-segment model, which is our model of choice throughout the paper.\(^3\)

In terms of the results, the recovered preference coefficients are in line with general expectations, in that consumers prefer vehicles being sold at lower prices, with lower mileage, fewer issues found during inspection, newer, with fewer past accidents, and fewer past owners. We also identify two different consumer segments, the first of which prioritizes evaluating vehicle photos, which produces relatively subjective information, whereas the second gives priority to more objective data, such as the vehicles’ inspection reports.

In terms of the results of the counterfactual analyses, we find that the way in which information is provided to consumers has statistically significant implications on conversion rates. Different information configurations can reduce conversion rates by \(-0.39\%\), or increase them by up to \(+1.65\%\), in relative terms. These figures are not only in line with what one would expect for the big-ticket items present in our context, but also with the intuition we provide in an illustrative example, which shows that information disclosure can affect the information sets of a pocket of relatively-indifferent consumers, who prefer to convert rather than acquire additional information.\(^4\) Notwithstanding this, we consider a benchmark scenario of ‘perfect information’, which reveals that the seller could attain a 0.51\% conversion rate increase \((10.41\%\) in relative terms\)) if it was able to provide all information to consumers which surpasses current computation capabilities. Additional work is required to appropriately extract heterogeneity from search data, and in fact, some work in progress on this topic is ongoing (see Morozov (2019)).

\(^3\)Details on model comparisons are available from the authors.

\(^4\)The effects of different information scenarios can also be estimated through the use of experiments. However, sellers are sometimes conservative in terms of providing different users with different experiences and moreover, each experimental condition requires specific design investments. Our modeling approach can easily compute the effects of multiple information scenarios (we report the findings of eleven scenarios in this paper). Given that experimentation is the gold standard of causal inference, one can also consider our estimates as starting points to isolate specific conditions of interest, to be tested experimentally at a later stage.
instantaneously.’ This scenario serves only as a theoretical benchmark, because moving all information to the consumers’ ex ante information set has other implications in terms of design, costs of thinking, information overload, etc.5

Re-estimation of the model under the knowledge gradient assumption allows us to evaluate the performance of this myopic heuristic, compared with that of the forward-looking approach. The knowledge gradient assumption has been shown to perform nearly- or exactly-optimally in theoretical analyses under certain conditions (e.g., Frazier, Powell, and Dayanik (2009) and Liang, Mu, and Syrgkanis (2017)), and has been documented to perform well across diverse empirical learning contexts (see Powell (2010)). Other search-related papers have similarly employed heuristic algorithms, sometimes suggesting that these may be more realistic than considering fully forward-looking consumers (e.g., Hodgson and Lewis (2017) and Dickstein (2018)). Despite having the same number of parameters, we find that the knowledge gradient method, in which consumers foresee only their immediate option value while considering whether to search or not, performs poorly in fitting the moments in the data when compared with the forward-looking model we propose.

The seminal methods and insights of consumer search have been established by both theoretical and empirical literature streams. Montgomery, Li, Srinivasan, and Liechty (2004) employ a dynamic multinomial probit model to characterize consumer online browsing for books, and find that first-order Markov approximations perform poorly in explaining the patterns in their dataset. Their findings highlight the need to consider at least second-order moments during estimation of search processes. The results by Koulayev (2013) alleviate this result, by proposing a Dirichlet-based model in which partial moment data can still be useful to estimate search models. Kim, Albuquerque, and Bronnenberg (2010) develop an empirical application of search based on Weitzman (1979), and use camcorder browsing behavior to estimate the search model. They also use the model to conduct welfare counterfactual analysis related to the availability of recommendation systems. Branco, Sun, and Villas-Boas (2012) characterize a search setting in which consumers can acquire information gradually, across multiple characteristics. Their model also takes into account the fact that the seller may choose price strategically to influence consumer search. More recently, Ke, Shen, and Villas-Boas (2016) characterize continuous-time search policies when consumers

5 See for example, Shugan (1980) and Hauser and Wernerfelt (1990).
are faced with multiple products.\footnote{See also Ke and Villas-Boas (2018) for a complete characterization of search across multiple alternatives, where search can be decreasingly informative.} Finally, Kim, Albuquerque, and Bronnenberg (2016) derive a probit-based search estimation approach, of particular benefit for settings with large consideration sets.

Although not always mentioned, the insights and advances available in the search literature are quite related to those of the learning literature (see, for example, Erdem and Keane (1996); Erdem (1998); Erdem, Imai, and Keane (2003)), in which consumers make purchase decisions \emph{while} they are learning about product characteristics/benefits. The difference is that, in that case, search is not geared towards a unique purchase, but it is performed concurrently with purchase decisions. See also Dickstein (2018) for a recent example of learning whilst consuming.

Ongoing research sharing some of the features in our paper exists. Ursu, Wang, and Chintagunta (2018) model piecemeal search behaviors by using search duration as a proxy for search intensity over different alternatives. The authors implement an estimator based on the work of Chick and Frazier (2012), who derive the optimal continuous time search policy for information search of one alternative, and derive a heuristic-based extension for the case of multiple alternatives. Choi, Dai, and Kim (2018) use the Weitzman search framework to derive the oligopoly equilibrium, and show that markets with lower search costs induce higher equilibrium prices. Chung, Chintagunta, and Misra (2019) rely on the same search framework to recover the market structure of the US automobile market, while allowing for the ex-ante and search characteristics to be correlated.

There has been a productive effort in describing consumer search patterns, through both descriptive and model-based methods. Johnson, Moe, Fader, Bellman, and Lohse (2004) document relatively short search sequences on websites selling books, c.d.’s and travel products, and also that consumer experience may lead to less search over time. Bronnenberg, Kim, and Mela (2016) analyze consumer search behavior for cameras, and find that consumers tend to focus on a small set of attributes, while also documenting state-dependence across searched attributes. Their findings also speak to those by Johnson, Moe, Fader, Bellman, and Lohse (2004), in that the chosen alternative is often first visited near the end of the search sequence. By allowing for arbitrary consumer search paths, our model can rationalize the state-dependence found in this literature, as well as the narrow consumer focus on
a few attributes. On the model-based front, research on the identification across specific search models includes De los Santos, Hortaçsu, and Wildenbeest (2012) and Honka and Chintagunta (2016), both of which support the fixed-sample search strategy, in relation to sequential search à la Weitzman.

Consumer search patterns can also be used to recover fundamental parameters. By adding formal structure to the search processes, or by considering additional sources of data, the work by Hong and Shum (2006), Honka (2014), Koulayev (2014), De Los Santos (2016), and Seiler and Pinna (2017) provides search cost estimates for consumers browsing for a variety of products, sold online and offline.

There exist additional ways in which consumer search is relevant to firms. Seiler (2013) finds that 70% of consumers are unaware of the price of detergent, and that price promotions can increase consumers’ incentives to search. Also in a physical retail setting, Elberg, Gardete, Macera, and Noton (2017) use search behaviors to rationalize consumers’ sensitivities to future deals, as induced by present ones. They go on to propose a slippery slope induced by offering promotions, because current deals may encourage future search and increase future price elasticities.

Search behaviors have also been characterized in situations where users have access to ‘refinements’: sorting and filtering tools that allow searching for alternatives more efficiently. Chen and Yao (2016) find that refinement tools induce more search, have positive effects on consumer utility, and lead to less concentrated market structures. De Los Santos and Koulayev (2017) propose a method to design product rankings to be used by consumers who have access to search refinements.

A standard assumption held in most of the work in the search literature is that consumers face a set of uncorrelated characteristics/alternatives. This assumption alleviates the estimation burden, because information acquisition no longer requires completely recomputing the search problem consumers face at each information set. Some exceptions include the work by Adam (2001), who considers the problem of an agent who faces a set of alternatives that may be correlated in an ex ante unknown way. In this case, searching an alternative yields information not only about that same alternative, but potentially also about some of the remaining ones. Santos, Hortaçsu, and Wildenbeest (2017) allow consumers to learn about their own utility distribution, and recover search cost bounds that can rationalize the observed online browsing and purchasing behavior. Hodgson and Lewis (2017) allow the in-
spection of a product to be informative about the utility that may be derived from inspecting others, and document that, because of learning, an initial bad product shown to a consumer can lead to ending search earlier. Our model also allows for learning, which follows from allowing for multiple observable characteristics, which may be related in arbitrary ways, as captured by their empirical joint distribution. Moreover, our model incorporates search over an unobservable characteristic, itself also potentially related to the observable ones. In our case, the information correlation structure induces learning, albeit in a different way than in the papers cited above. In our case, learning about a characteristic of an alternative is informative about the remaining characteristics of that alternative (within-alternative correlation); whereas in the literature above, uncertainty about one’s own preferences induces cross-alternative correlations. Our model applies to contexts where consumers already understand their preferences well, and are mostly uncertain about different product attribute levels, whereas the assumptions in Santos, Hortaçsu, and Wildenbeest (2017) and Hodgson and Lewis (2017) apply to situations where consumers are first and foremost uncertain about their own preferences, with incremental information acquisition being costless along characteristics of each alternative.\footnote{See also Kamenica (2008); Guo and Zhang (2012); Cao and Zhang (2017); Gardete and Guo (2018) for settings in which consumers learn about their own preferences, or about the fit between their preferences with the product characteristics.}

The next section describes our context and dataset. Section 3 describes the search model as well as the empirical analysis, including consumer beliefs, identification and estimation details. Sections 4 and 5 present the empirical results and the counterfactual analyses, respectively, and Section 6 concludes.

2 Data

2.1 Browsing Behavior

Our dataset comprises all online browsing activity on the website of a used car seller, Shift, between February and September 2016.\footnote{Firm’s website: www.shift.com.} The firm operates in a number of geographic markets, listing more than 4,000 vehicles during the sample period. Upon arrival to the website, users can click through a number of filters in order to focus on the vehicles of interest. The website lists a number of vehicle details. First, on the main listing page (see
Figure 1 for screenshot of main listing page, applicable to the data collection period), each vehicle photo is accompanied by make-model information, year, price, and mileage data. Additional information is available in each vehicle’s detail page (see annotated screenshot in Figure 2, applicable to the data collection period). In addition, users can 1) browse through pictures, 2) access the report of the dealer’s vehicle inspection, and 3) access the vehicle’s history, which includes additional information such as the number of previous owners and the number of accidents the vehicle has been involved in. Accessing each of these informative elements (pictures, inspection reports, and vehicle histories) requires deliberate action by consumers through clicks. The dataset collects all such browsing information, including all user sessions, webpage visits, and click actions. Although vehicles are not sold directly online, the dealer uses a differentiated business model that provides a primary conversion variable. Upon settling on a vehicle, consumers can order a test drive by filling out a form with the relevant information. The company then follows up to confirm the time and location of the test drive, usually taking place near a location convenient to the customer. Successful test drive appointments are available in the dataset and are used to characterize consumer conversion. Given the broad product line offered by the firm, we focus on the search data related to browsing activity of sedan vehicles.

Table 1 presents descriptive statistics of the dataset, comprising information about the search behavior of the 24,116 users on the website. The descriptive statistics point to two main patterns. First, search activity can take place over long periods of time. For example, although each user browsed through an average of 2.28 vehicle profile pages, it is possible to find users who browse tens or hundreds of vehicles. This is expected, given the complexity of the products involved as well as their cost. Search activity is divided in browsing sessions, 

---

9 At the time of our sample, all information-disclosing actions involved either clicks, when accessed via computer, or taps and swipes, when accessed through mobile and tablet devices. We restrict ourselves to browsing sessions via computer platforms since mobile and tablet swipes were found not to be consistently captured in the dataset. Moreover, we also eliminated consumers whose modal device was a phone or a tablet. We also conduct several data cleaning activities in order to eliminate ‘bouncing’ behavior and other noisy online activity, as we describe in Appendix A.

10 Also, the firm provides information on the vehicle prices on its website, which cannot be bargained over. Given the high conversion rates, we only model consumer behavior up to the first conversion. Related to this, we also eliminate redundant search decisions for objective information. We believe these are related to consumer memory. Although these could be incorporated in the model, they are not the main focus of the analysis. The working assumption for the counterfactual analysis is that consumers will always conduct the necessary searches to refresh their memories, if needed.

11 The sedan category is the largest one. Search behavior changes significantly depending on the focal category. For example, the sports cars category attracts a high number of users, but displays lower conversion/visit ratios.
defined as sets of events taking place within sequential intervals of 30 minutes. On the high end, a few consumers went through more than 700 sessions, spread across three quarters of a year.

The other insight taken from the descriptive numbers is that search behaviors can vary wildly, per inspection of the ‘Min’ and ‘Max’ columns of Table 1. While some users did not browse more than 1 vehicle, some browsed more than 100. The remaining statistics also indicate significant reasonable variation in search patterns.

One of the key dependent variables in the dataset is whether consumers end up converting. It is natural to expect that consumers who order test drives behave differently during their search than the ones who do not. Figure 3 depicts the kernel densities of average session times across converters and non-converters. Across a relatively long tail, we find converters engage in longer browsing sessions, whereas non-converters concentrate their activities around sessions with durations below 15 minutes.

The model captures the heterogeneity in search patterns in two ways. First, we restrict each consumer’s consideration set to the set of vehicles that was ever accessed by each individual. This means that any search action surrounding a vehicle in the data implies that the vehicle belongs to the consumer’s consideration set. This assumption allows us to abstract from the specific filters and the sorting tool consumers may have used to initially discover the vehicles they are interested in. While allowing for different consideration sets across consumers reduces the computational burden of computing the value function, it implies solving a large number of smaller value functions, one for each consumer.\footnote{We refer the reader to the Introduction section for a detailed explanation of the advantages and disadvantages of introducing heterogeneous consideration sets, in light of the current literature.} We focus on the 89% of consumers who evaluated up to four vehicles in the data, yielding a working sample with 12,887 consumers (91% browse up to five vehicles).

The second way the model accounts for heterogeneity in search patterns is by incorporating a latent class distribution over search cost levels, such that some consumers may be willing to search certain aspects more than others. Moreover, each segment is allowed to have different search costs related to each information-gathering activity.

Finally, idiosyncratic shocks can also explain differences in search activities across consumers.
2.2 Vehicle Characteristics

We now turn our attention to the heterogenous characteristic space of the 1,573 sedan vehicles in the dataset. Each vehicle is described by its make, model, color, price, mileage, number of inspection notes, age, number of accidents, and number of previous owners.

Figure 4 depicts the histograms and cross-scatter plots of vehicle characteristics, excluding the categorical variables make, model, and color. The main diagonal reveals that vehicles are relatively heterogeneous in terms of their characteristics. Moreover, characteristics follow very different marginal distributions. For example, price and mileage are continuous variables and exhibit non-normal distributions. The remaining characteristics are countable, with seemingly different distributional properties as well: The number of previous owners starts at 1 and is strictly decreasing, whereas the number of inspection notes and the vehicle’s age (in years) are non-monotonic.

Table 2 depicts the correlation matrix across the ordinal characteristics. With the exception of price, all characteristics are pairwise positively correlated, which is not surprising since they are likely to be negatively associated with the vehicle value. The negative correlation between price and these characteristics is also expected: It is an indication that the seller prices the vehicles according to the appeal of the characteristics. While the actual relationships across variables are non-linear, the results in Table 2 are in line with Figure 4. These analyses stress the need to allow for flexible marginal distributions of the vehicle characteristics as well as to allow for flexible relationships among them. As we discuss later, we introduce a flexible method to fit the multivariate distribution of the vehicles’ characteristics, while including the additional effects of make, model, and color.

2.3 Preference Weights

In order to briefly characterize consumer preferences, we regress consumers’ conversions on vehicle characteristics by use of a logit model. Table 3 summarizes the results. As expected, consumers dislike increments to all vehicle characteristics. All results are highly significant, except for the number of previous accidents a vehicle suffered in the past, whose parameter is only significant at the 10% level.

The second column presents the marginal effects measured in odds ratios, which allows

\[13\text{ Inspection reports feature one inspection note for each issue found by the dealer.} \]
for a clear interpretation of the relative magnitude of the results. Keeping the remaining characteristics constant, the model implies that a price increase of $10,000 makes a vehicle’s purchase probability decrease to approximately a fourth of its initial value, and an increase in 100,000 miles translates into the probability of a sale decreasing approximately to a third.

While the results are economically significant, they translate into relatively low elasticities. There may be two underlying reasons. First, an estimated low elasticity may be partially induced by the fact that some consumers would not have been actually exposed to all characteristic changes, due to search frictions. The logit model assumes perfect information, however, and so explains the low response through low elasticity estimates. The second reason affecting these estimates is the impact of unobserved utility factors, which may be correlated with the observed vehicle characteristics. We later detail that considering consumers’ search behaviors allows us to incorporate the correlations between unobserved utility that consumers search over with the observable vehicle characteristics. For example, a positive correlation between price and the unobserved component will likely bias price sensitivity estimates toward zero. The search model can characterize the propensity of search for the unobserved utility component conditional on vehicles’ price levels and remaining characteristics. In this way, search behavior is used to recover consistent estimates, through estimation of the correlations between the observable components and the unobserved one.

3 Model

3.1 Utility under Perfect Information

As in most discrete choice applications, we assume consumers derive linear utility from the characteristics of the alternatives. Under perfect information, consumer $i$’s indirect utility for purchasing vehicle $j$ is equal to

$$v_{ijt} = \sum_{k=1}^{K} \beta_k x_{jk} + v_{ij} + \epsilon_{ijt}$$  \hspace{1cm} (1)$$

and earns utility

$$v_{i0t} = \epsilon_{i0t}$$  \hspace{1cm} (2)$$
if she decides for the outside option. Above, $j$ indexes vehicles and $k$ indexes observable (by the researchers) vehicle characteristics. Component $\nu_{ij}$ is an idiosyncratic preference for vehicle $j$, and $\epsilon_{ijt}$ is a preference shock, learned by the consumer at time $t$. Both terms $\nu_{ij}$ and $\epsilon_{ijt}$ are unknown to the econometricians, and the preference shock $\epsilon_{ijt}$ is uncorrelated across consumers, vehicles and time. Except for $\epsilon_{ijt}$, all vehicle characteristics may be correlated. Characteristics $\{x_{jk}\}$ are constant across consumers for the same vehicle (e.g., price, mileage, etc), and may be correlated with each other. For example, as discussed in the descriptive analysis, it is reasonable to expect an older vehicle to have higher mileage as well. These characteristics may also be correlated with the idiosyncratic characteristic $\nu_{ij}$ which, unlike $\{x_{jk}\}$, may yield different realizations across consumers. The introduction of term $\nu_{ij}$ captures the fact that, during search, consumers may learn more than the objective characteristics that the econometricians are able to observe directly in the dataset. Including term $\nu_{ij}$ takes into account that a search action (like photo browsing) may reveal subjective information that is not only challenging to quantify, but also whose judgment may vary across users.\textsuperscript{14}

The website reveals some vehicle characteristics on its main listing page, including a picture, price, mileage, age, and make/model information. We include these characteristics in the consumers’ initial information set. Because these specific characteristics are extremely easy to observe, search for information is modeled over and above these elements, as we now explain.

Table 4 organizes the vehicle characteristics and presents the search action correspondence that allows users to discover specific characteristics. We assume consumers know their own preference parameters, as well as the vehicle characteristics included in the initial information set. Users can learn additional vehicle information by taking search actions with respect to each vehicle. Inspection of a vehicle’s history reveals both the number of previous owners of the vehicle and the number of accidents the vehicle was involved in. Accessing inspection reports reveals information about the number of issues identified during the inspection of the vehicle by the seller. Finally, browsing through vehicle photos informs users of the utility component $\nu_{ij}$.\textsuperscript{15}

\textsuperscript{14}One can also decompose $\nu_{ij}$ as $\beta_{i},x_{ij}'$, where $x_{ij}'$ is an unobserved product characteristic. We keep notation $\nu_{ij}$, given the impossibility of estimating the distribution of preferences separately from the level of the unknown product characteristic $x_{ij}'$.

\textsuperscript{15}We discretize the photo-browsing activity because of the following rationale. Theoretically, the act of
3.2 Search Dynamics

During their searches, consumers decide whether to take subsequent information-acquisition actions about vehicles in their consideration sets, or to terminate search either by ordering a test drive or opting for the outside option. Given the discussion in the previous section, it is clear that search behaviors are cumulative processes, taking into account the information already available at each decision point. For example, a customer may stop or continue searching a vehicle depending on how many inspection issues she finds. Moreover, because features are correlated, a user may expect a vehicle to have had more owners after observing its high mileage. Clearly, expectations of unknown features should be conditioned on the known ones.

**State Variables.** We characterize the search problem of a consumer who acts ex-ante optimally according to her beliefs. Consumer \( i \) considers alternatives \( 1 \ldots J_i \), each with \( k \) characteristics, as described before. All consumers also have access to an outside option, whose deterministic utility is normalized to zero. Consumers have \( S \) search actions available to them for each alternative, each of which maps into learning one or more characteristics. As before, there exist shocks \( \epsilon_{ijt} \) that affect each vehicle’s utility in an idiosyncratic fashion. Moving towards a Bellman equation framework, we omit time subscripts, and assume the information-acquisition actions are also perturbed by preference shocks, denoted as \( \epsilon_{ij}^s \), where \( s \in S \) denotes a search action. These preference shocks are learned contemporaneously by consumers, and represent unobservable influences that may randomly affect consumers’ actions. We denote all of user \( i \)’s preference shocks as \( \epsilon_i \).

We summarize the characteristics known by consumer \( i \) about vehicle \( j \) by set \( \Omega_{ij} \), and denote the collection of all information known about all vehicles in consumer \( i \)’s consideration set by \( \Omega_i \). For example, before initiating search for specific vehicles, user \( i \) is endowed with information \( \Omega_{ij} = \{\text{make}_j, \text{model}_j, \text{price}_j, \text{mileage}_j, \text{age}_j, \text{color}_j\} \) about vehicle \( j \). The information set is then expanded according to the consumer’s search decisions.

---

browsing each photo adds information to the consumer’s posterior utility about the focal vehicle. In practice, the resulting search sequences can become extremely long, increasing computation requirements to attain precise calculation of the likelihood function.
**Decision-Making.** At each decision point, consumer $i$ solves the problem:

$$
V^i(\Omega_i, \epsilon_i) = \max \left\{ \begin{array}{ll}
V^i_0(\Omega_i, \epsilon_{i,1}), & \vdots, V^i_{J_i}(\Omega_i, \epsilon_{i,J_i}) \end{array} \right\}
$$

where $V^i(\cdot)$ is the value function for consumer $i$. At each decision point, the consumer may decide to stop her search and make a final selection, or continue learning by taking a search action. $V^i_j(\cdot)$ denotes the expected value of stopping search and selecting one of the $J_i + 1$ alternatives, and $V^i_{j,s}(\cdot)$ denotes the continuation value of taking a search action $s$ w.r.t. vehicle $j$.

If the consumer stops her search, she earns expected utility

$$
V^i_j(\Omega_i, \epsilon_i) = E(v_{ij} \mid \Omega_i, \epsilon_i)
$$

conditional on the information available. If instead she decides to take a search action, say action $s$ w.r.t. vehicle $j$, she expects continuation utility

$$
V^i_{j,s}(\Omega_i, \epsilon_i) = -c_s + E_{\epsilon'_i, \omega_{js}} (V(\Omega_i \cup \omega_{js}, \epsilon'_i) \mid \Omega_i) + \epsilon_{ij}
$$

where $j$ and $s$ denote the vehicle and search actions to be maximized over. Vector $\omega_{js}$ contains the characteristics of vehicle $j$ to be learned with action $s$. The correspondence between search actions and learned characteristics are presented in Tables 4 and 5.

We allow consumers to incur different search costs depending on their actions, namely at rates $c_s$, $s \in \{1, 2, 3\}$. We later explain that we allow $c_s$ to be heterogeneous across consumers as well.

Expression $\{\Omega_i \cup \omega_{js}\}$ captures the augmented information set that the consumer will
have access to, if she decides on search action $s$ for vehicle $j$. For example, if she starts out by searching vehicle $j$’s photos, her information set about vehicle $j$ will transition from

$$\Omega_{ij} = \{\text{make}_j, \text{model}_j, \text{price}_j, \text{mileage}_j, \text{age}_j, \text{color}_j\}$$

(7)

to

$$\Omega_{ij}’ = \{\text{make}_j, \text{model}_j, \text{price}_j, \text{mileage}_j, \text{age}_j, \text{color}_j, \nu_{ij}\}$$

(8)

Consumer $i$’s decision to search incorporates the fact that she already has some information about vehicle $j$. Hence, the expectation operator in expression (6) is taken with respect to the preference shocks $\epsilon’_i$ as well as to the information $\omega_{js}$ that the consumer will obtain through search, and which is affected by the information in $\Omega_i$.

### 3.3 Beliefs

We assume consumers hold beliefs consistent with the distribution of characteristics observed in the data. Let $X = [X_1..X_K]$ be the random vector of vehicle characteristics with realizations observable both to the consumer and to the researchers (i.e., all characteristics excluding $\nu_{ij}$ and $\epsilon_{ijt}$). We denote the multivariate cumulative distribution (c.d.f.) of these characteristics by

$$G_{\text{make}_\text{model}, \text{price}, \text{mileage}, \text{age}, \text{color}, \text{owners}, \text{accidents}, \text{notes}}$$

(9)

In order to estimate this joint distribution while preserving the different natures of the marginal distributions as well as the cross-correlation patterns, we employ a statistical copula. Sklar (1959) showed that all multivariate c.d.f.’s $F_{X_1..X_K}$ admit copula representations

$$C(F_{X_1}(x_1),...,F_{X_K}(x_K))$$

(10)

where each $F_{X_k}(\cdot)$ is the marginal c.d.f. of variable $X_k$. Taking advantage of Sklar’s result is, however, complicated by two factors. First, variables such as color and make-model information are categorical r.v.’s. Given their lack of ordinality, cross-correlations are relatively meaningless for these variables. To see this, suppose $X_1 \leq 0.2$ were to define a vehicle as yellow, and $X_1 \in (0.2,0.3)$ meant that the vehicle was white. It follows that proper estimation of a copula parameter between r.v.’s $X_1$ and $X_2$ ($X_2$ could represent mileage, for example)
would require an unrealistic degree of flexibility in order to capture the lack of ordinality in \( X_1 \). The second challenge is that the unobserved preference shock \( \nu_{ij} \) may also be correlated with the observable characteristics, and so needs to be incorporated into this framework.

We start by addressing the first issue. Estimating the multivariate distribution with categorical variables could, in principle, be implemented by estimating a separate distribution for each level of the variables \{make\_model, color\}. However, this method can be cumbersome to work with, and imposes large data requirements. Because copula estimation can rely on parametric or non-parametric estimation of the marginal distributions, we employ a hybrid approach: We estimate the marginal distribution of the ordinal variables, net of a parameterized effect of the non-ordinal variables. We employ separate ordered probit regressions to each of the ordinal variables using the categorical variables as regressors. For example, consider the case of the number of inspection notes found for a vehicle, which may have one of seven values in the data (i.e., zero to six issues found during inspection). We estimate an ordered probit regression for inspection notes according to the model

\[
\text{notes}_j = \begin{cases} 
0, & y^*_j \leq \mu_0 \\
1, & \mu_0 < y^*_j \leq \mu_1 \\
\vdots & \vdots \\
6, & y^*_j > \mu_6 
\end{cases}
\]

(11)

where

\[
y^*_j = \beta_{\text{notes}}^\text{make\_model}_j + \beta_{\text{notes}}^\text{color}_j + \epsilon_{\text{notes}}^j
\]

(12)

and make\_model and color are indicator variables, and \( \epsilon_{\text{notes}}^j \sim N(0, 1) \). We employ the same approach to the remaining variables, including price and mileage, which we discretize into five levels each.\(^{17}\) Our approach assumes a functional form for the relationship between the categorical and ordinal variables. At the same time, it is fully flexible in terms of the distribution of the ordinal variables, as it associates a probability mass to each characteristic level. Figure 5 depicts an example of the cutoffs recovered for the case of the inspection notes.

The result of the ordered probit regressions is a set of estimates for the make-model

\(^{17}\)Discretizing these variables into additional levels is straightforward, with a low penalty for our estimator.
and color fixed effects. Note that the estimated residuals $\hat{\epsilon}_j$ are independent of the non-ordinal variables. Because each discrete value of the dependent variables can be generated by several values of $\epsilon_j$, we need to define a specification for estimator $\hat{\epsilon}_j$. We use the maximum likelihood estimator of $\hat{\epsilon}_j$, so that each level of each characteristic is associated with a unique level of the estimated residual $\hat{\epsilon}_j$. By construction, these residuals are independent of the categorical variables make model and color.$^{18}$ We denote the joint distribution of the estimated residuals as

$$F_{\text{price, mileage, age, owners, accidents, notes, make model, color}}$$

When the residuals above are evaluated at their empirical c.d.f.’s, they induce random variables $u = [u_1...u_K]'$, which are uniformly distributed by construction. These variables are introduced into a Gaussian copula:

$$c_\Sigma(u) = \frac{1}{\sqrt{\det \Sigma}} \exp \left\{ -\frac{1}{2} \left( \begin{array}{c} \Phi^{-1}(u_1) \\ \vdots \\ \Phi^{-1}(u_K) \end{array} \right)' \Sigma^{-1} \left( \begin{array}{c} \Phi^{-1}(u_1) \\ \vdots \\ \Phi^{-1}(u_K) \end{array} \right) \right\}$$

where

$$u = F(\epsilon) = \left( \begin{array}{c} F_1(\epsilon^1) \\ \vdots \\ F_K(\epsilon^K) \end{array} \right)$$

and where matrix $\Sigma$ is a correlation matrix (its main diagonal equals one, by construction), which moderates the relationships between the residuals. Function $\Phi^{-1}(\cdot)$ is the inverse c.d.f. of the standardized normal, and functions $F_k$’s are the empirical marginal distributions of each characteristic. Note that $c_\Sigma(u)$ can be interpreted as a joint normal probability density function (p.d.f.) of random vector $(\Phi^{-1}(u_1) ... \Phi^{-1}(u_K))'$. We simplify the notation by defining $z_k = \Phi^{-1}(u_k)$, so that the copula can be readily interpreted as a p.d.f. of the

$^{18}$An alternative (and more demanding approach for our estimation strategy) would be to associate multiple levels of residuals to each characteristic level, by simulating $\epsilon_j$’s from intervals of the ordinal probit regressions.
random vector $Z = \{z_1..z_K\}$:

$$f_{Z_{1..K}}(z) = \frac{1}{\sqrt{\det \Sigma}} \exp \left\{ -\frac{1}{2} \left( \begin{array}{c} z_1 \\ \vdots \\ z_K \end{array} \right) \cdot \Sigma^{-1} \cdot \left( \begin{array}{c} z_1 \\ \vdots \\ z_K \end{array} \right) \right\}$$ (15)

where $z_k \sim N(0, 1)$, by construction.

Above, the elements of random vector $Z$ follow non-independent standard normal distributions, with covariance matrix equal to $\Sigma$.

Estimation of the copula relies on the steps summarized in Figure A:

Figure A: Copula Estimation and the Marginal p.d.f. of $z_k$

$$G_k^{-1}(x_k) \xrightarrow{\varepsilon_k} F_k(\varepsilon_k) \xrightarrow{u_k} \Phi^{-1}(u_k) \xrightarrow{z_k} f_{Z_k}$$

First, the residuals of each characteristic $\varepsilon_k$, orthogonal to make-model and color information, are estimated through the transformation

$$\varepsilon_k = G_k^{-1}(y_k, X)$$ (16)

where $G_k^{-1}$ is the maximum likelihood estimate of the errors of the ordinal probit regression for each characteristic, $y_k$ is a vector with the levels of characteristic $k$, and $X$ is a set of indicator variables with make-model and color information for each vehicle. Calculating the empirical percentiles of $\varepsilon_k$ for each $k$ (or in different words, evaluating the empirical distributions $F_k$ at values $\varepsilon_k$), yields uniformly-distributed random variables $u_k = F_k(\varepsilon_k)$. Then, applying the inverse standard normal c.d.f. yields normally-distributed random variables $z_k = \Phi^{-1}(u_k)$. Finally, $\Sigma$ is easily obtained by calculating the empirical correlation matrix of $(Z_1..Z_K)'$. The collection of steps described above allows us to characterize the relationships across variables in a flexible way, net of the categorical variables, despite the different marginal distributions. As we explain later, the process depicted in Figure A is also used in
reverse during estimation.

It remains to solve the second estimation challenge, which is to include the unobserved utility component $\nu_{ij}$ to this framework. We assume this component follows a normal distribution, such that it can be naturally added to the statistical copula by expanding the ‘multivariate density’ appropriately:

$$f_{z_1 \ldots z_K, \nu}(z) = \frac{1}{\sqrt{\text{det} \Sigma_0}} \exp \left\{ -\frac{1}{2} \begin{pmatrix} z_1 \\ \vdots \\ z_K \\ \nu \end{pmatrix}^{\prime} \begin{pmatrix} \Sigma & \Sigma_{z\nu} \\ \Sigma_{\nu z} & \sigma_\nu^2 \end{pmatrix}^{-1} \begin{pmatrix} z_1 \\ \vdots \\ z_K \\ \nu \end{pmatrix} \right\}$$

(17)

The covariance matrix $\Sigma$ is now bordered by the relationships between the unobserved component $\nu$ and each of the observable attributes, making up matrix $\Sigma_0$. Vector $\sigma_{z\nu}$ contains the covariances between $\nu$ and each ‘normalized characteristic’ $z_k$, and $\sigma_\nu^2$ is a scalar corresponding to the variance of $\nu$. This specification allows us to identify correlations between observed characteristics and the unobserved utility component.

### 3.4 Estimation

Search models are notably complex to estimate. Even the relatively simple Weitzman-based search model often requires the use of complex simulation-based estimation. Our setting is more complicated, due to the fact that we allow consumers to search alternatives in a piecemeal fashion while allowing for correlated characteristics. In our case, the Weitzman conditions that support the optimality of the parsimonious “Pandora search rule” are not applicable, and so we are required to characterize the full dynamic problem faced by consumers.

A resulting challenge is the large size of the state space, part of which is unobservable. In addition to ex ante knowledge, each consumer can take up to three search actions per vehicle, and learn that its characteristics may be one of seven levels in terms of the number of inspection notes, one of four levels in terms of number of accidents, one of six levels in terms of number of past owners, and one of $\omega$ discrete levels used to simulate the unobserved characteristic. As a result, a vehicle’s information in a consumer’s state space has the
following number of elements:

\[ |\Omega_{ij}| = (1 + 7) \times (1 + 4 \times 6) \times (1 + \omega) = 88(1 + w) \]  

(18)

where the ‘(1 + •)’ structure above takes into account that consumers may not take all search actions. When \( J_i \) is the number of vehicles in a consumer’s consideration set, the size of the state space related to information acquisition for consumer \( i \) equals

\[ |\Omega_i| = J_i^{88(1+w)} \]  

(19)

For example, a consumer with four vehicles in her consideration set, and with unobserved characteristics approximated with \( w = 5 \) points, will face approximately \( 7.72 \times 10^{317} \) possible states. In addition to this, our context presents the challenge that we are required to solve as many value functions as the number of consumers in the sample, since we allow consumers to hold different consideration sets (consumer consideration sets rarely overlap exactly).

In order to estimate our model, we develop an estimator that relies on what we call ‘state bootstrapping.’ The approach produces an estimator that draws from the state space according to the likelihood of each realization, as implied by the current parameter guess.

Our approach has a number of advantages. First, it is scalable: it is possible to represent and estimate a large dynamic program in a modern computer. Second, the resulting decision tree is “small” whenever the search context is informationally path-independent, e.g., when the order in which search actions A and B are taken is irrelevant for the resulting information state. Third, the estimation procedure takes into account that the distribution over characteristics may itself be unknown (i.e., the covariance terms between the unobserved and observed characteristics), by incorporating a sparse support for the distribution of unknown characteristics (based on the work of Heiss and Winschel (2008)), whose probabilities change as a function of the guess for the covariance parameters. This approach increases the number of distribution support points for the unobserved utility component automatically, as the number of alternatives increases.

We refer the reader to Appendix C for the precise estimator details. It is especially useful for search problems, in which possibilities in the state space are eliminated as the decision-maker searches. We also present a sketch of the pseudocode necessary to represent
the corresponding value function in the appendix. We now provide an outline of the estimation procedure.

**First Stage Estimation Steps**

- First, estimate matrix $\Sigma$ (correlation of observable characteristics), according to the steps laid out in Section 3.3. (Note that the additional border that makes up matrix $\Sigma_0$ is estimated later, with the remaining model parameters).

- Set the number of simulations $R$. For each consumer, simulate $R$ sets of characteristics for each vehicle in her consideration set, conditional on the vehicle’s ex-ante characteristics. Refer to Figure B for the steps involved in this simulation, discussed later.\(^{19}\)

- Save both the simulated vehicle characteristics as well as the simulated $z_k$ draws (see Figure A), as a later input for the model estimation.

**Second Stage Estimation Steps**

- Select the initial values for the preference parameters, the search cost parameters, and the covariance parameters $\sigma_{z\nu}$. (We later explain that $\sigma_{\nu}^2$ is normalized to one.)

- If there exist new guesses for the covariance parameters, redraw values of $\nu$ conditional on the remaining characteristics.

- For each consumer, create or update the decision tree by first populating the utilities under perfect information, and then ‘undoing’ search decisions and building the tree backward. (This is a complex step, since the decision tree holds information about beliefs, terminal decisions, and transition probabilities, which partially depend on the simulations taken before estimation. Please refer to the appendix for the construction and updating steps of the decision tree).

\(^{19}\)In appendix C.5, we discuss the point that our estimator includes as one of the simulations, the search characteristics of the actual vehicles considered by the consumer in the dataset. Including that unique simulation draw does not affect estimator consistency.
• Once created/updated, use the decision tree to calculate the joint probability of the consumer’s search decisions observed in the data, for each simulation of the unobserved component.

• Average the probabilities across simulations and sum them across consumers. The result is the log-likelihood evaluation for the sample. Decide convergence and/or attempt new parameter estimates.

The key steps in our estimator are related to the construction of the decision tree, as a function of the simulated draws. Rather than keeping track of the whole joint distribution of characteristics, our estimator derives the decision problem as a function of the simulated draws of search characteristics as well as of the unobserved utility component. In other words, the estimator scales the decision problem as a function of the drawn simulations.

The steps above are laid out very concisely, especially in terms of describing the construction of the decision tree. For expositional purposes, here we provide an example of the decision tree induced by the case where the consumer faces only observed characteristics (by the researchers). We explain the inclusion of the unobserved characteristic in Appendix C. Each consumers’ dynamic problem is represented by a separate decision tree, with one ‘node’ for each possible configuration of her search actions. For example, a consumer with only one vehicle in her consideration set and three available search actions generates a tree with \(2^3\) nodes. Each node has a number of allowable search actions. For example, if the consumer is in node “010” (i.e., the binary representation for the consumer having taken the second search action), then actions “100” and “001” are still available. The decision of selecting a terminal decision, be it ordering a test drive or selecting the outside option, are available at all nodes as well. The result is that each node has a collection of allowed actions.

When the consumer arrives to node “010”, she may be in one of a number of information states. We call each state a “partition”, for reasons we explain further below. Finally, search actions transition consumers across nodes (e.g., action “100” leads consumer from node “010” - if she has already taken the second search action - to node “110”).

In order to clarify the decision tree further, we consider the case of a decision-maker (DM) who faces only one alternative with two characteristics, each of which she may decide to search or not. Assume that the support of the characteristics is \(x_1 \in \{1, 2\}\) and \(x_2 \in \{2, 3\}\), and that they are distributed according to some joint distribution. We depict the DM’s
decision tree in Figure 6. Before estimation, we take draws from the joint distribution of characteristics (we use four draws here, for illustration purposes only). Suppose that we obtain draws \{1, 2\}, \{1, 3\}, \{2, 2\}, and \{1, 3\}, as represented in the top-left corner of Figure 6, and the ‘freq.’ column denotes the frequency with which each simulation was drawn.

In the figure, the tree extends from right to left, starting at node “00” - i.e., where no search action has been taken yet. This node has a single information set \(P_1 = \{1, 2, 3\}\): Before search takes place, any of the simulations 1, 2, or 3 could correspond to the truth, and the consumer is required to search, in order to understand which case she is actually facing. For example, in the data, the true characteristics could be \(x_1 = 1\) and \(x_2 = 2\).

At node “00”, any of the simulations may represent the true vehicle characteristics, but the decision maker does not know which one. Suppose the consumer decided to take search action 01 (top arrow), which informs her of characteristic \(x_2\). She understands that she will transition from node 00 to node 01, and specifically, that her new information state will be either \{1, 3\} or \{2\}. To see this, note that because \(x_2 \in \{2, 3\}\), implies that learning it partitions the initial simulation space into two subsets. If the DM learns that \(x_2 = 2\), then she understands that the correct simulation is either draw 1 or 3, which indeed exhibits \(x_2 = 2\); otherwise, the correct simulation must be element 2 (the only element with \(x_2 = 3\) in this example).

The state bootstrapping approach is also useful to calculate transition probabilities. Specifically, because the simulations are drawn from the joint distribution of characteristics, the frequency column can be used to easily calculate transition probabilities. Taking action 01 at node 00 transitions the decision maker to partition \{1, 3\} with \(1/2\) probability, because simulations 1 and 3 were drawn once each, whereas simulation 2 was drawn twice. At node 00, the consumer may also take search action 10, and learn \(x_1\), which partitions the state space into partitions \{1, 2\} and \{3\} with transition probabilities \(3/4\) and \(1/4\), respectively.

We provide a depiction of a decision tree for a consumer searching over a single vehicle in Figure 7. The tree depicts the consumer’s possible 8 action states (i.e., combinations of search actions taken), as well as the probabilistic transitions across them. The decision tree is highly non-linear, and different sequences of search decisions can take the consumer to the same information state.

A few additional notes are in order. First, note that the number of nodes is easy to

\[20\text{See Appendix C.5 for additional details.} \]
calculate, as it is equal to the number of possible search actions configurations. Hence, a consumer with a four-vehicle consideration set and three search actions per vehicle induces a tree with \((2^3)^{J_i^4} = 4,096\) nodes. Second, the number of information partitions is ex-ante impossible to predict, since it depends on the distribution used to draw the simulations as well as on the actual draws. In the limit, if the distribution has very low variance, most nodes would have only a few partitions. In the other extreme, when the variance of the distribution is extremely high, then each action may split partitions in many ways, yielding a much larger number. Third, creating and storing a tree is relatively challenging (we present a code representation in section C). Because of the arbitrary behavior of partitions, which depends on the specific simulations drawn, the most convenient tree representation in common computer languages is through the use of pointers. While our approach precludes standard vectorization techniques, the cost comes at the great benefit that the calculation of expected utilities and of transition probabilities is relatively trivial, and the tree is stored efficiently.

**Simulation.** During the decision process, consumers need to integrate over the information they are likely to learn from different search actions, conditional on the information already available. These expectations are approximated via simulation. Below, we depict the process of drawing characteristic levels \(x_l\), given the levels of some other characteristics, \(x_k\). We have already defined the joint distribution of the random variables \(\{z_k, \nu\}\), underlying the vehicles’ characteristics (see equation (17)). However, except for \(\nu\), these variables do not enter utility functions directly. Additional steps are required to take draws of the characteristics they underlie, as depicted in Figure B. Taking draws of vehicle characteristics \(x_l\), conditional on characteristics \(x_k\), involves reversing the process outlined before. For example, suppose we intend to take draws of characteristic \textit{age}, conditional on \textit{price}. First, we convert \textit{price} to \(z_{\text{price}}\), by following the steps depicted on the top arrows of Figure B. Second, we use the conditional normal p.d.f. to take draws \(\{z'_{\text{age}}\}_{r=1}^R\), conditional on \(z_{\text{price}}\). We then carry out the reverse process, depicted on the bottom of Figure B, to produce the intended simulations. Before estimation, we draw simulations of all of the observable search characteristics, conditioning on all of the ex-ante characteristics. These draws are saved and available during the estimation of the model parameters. In order to save computation time, we also save the draws \(z^r\), which are used to simulate values \(\nu(r)\) during the model estimation. These draws
Figure B: Obtaining draws of characteristics $x_l$, conditional on characteristics $x_k$

\[
\begin{align*}
  G_k^{-1} & \quad F_k & \quad \Phi^{-1} \\
  x_k & \quad \varepsilon_k & \quad u_k & \quad z_k \\
  x_l^{(r)} & \quad \varepsilon_l^{(r)} & \quad u_l^{(r)} & \quad z_l^{(r)} \\
  G_k & \quad F_k^{-1} & \quad \Phi & \quad f_{z_l|z_k}
\end{align*}
\]

Note: The process above applies to observable characteristics. The component $\nu$ is assumed to follow a normal distribution, from which draws can be readily taken, given knowledge of covariance matrix $\Sigma$.

are updated for each parameter guess of $\sigma_{z\nu}$.

**Variance-Covariance Matrix.** Matrix $\Sigma_0$ moderates the relations between the vehicles’ observed characteristics, and its border includes the relations with the idiosyncratic utility component $\nu$. As we further discuss in the Identification section, we normalize parameter $\sigma^2_{\nu}$ to one. The remaining elements of the border of the variance-covariance matrix are estimated under the restriction that the resulting variance-covariance matrix is positive semi-definite (p.s.d.). First, we find an upper triangular matrix $C$ such that

\[
C' \cdot C = \begin{pmatrix}
\Sigma & \sigma_{z\nu} \\
\sigma_{z\nu}' & 1
\end{pmatrix}
\]  
(20)

The process of finding $C$ follows from common Cholesky decomposition. In appendix B, we explain how we parameterize matrix $C$ in order for $\Sigma_0$ to be simultaneously consistent with the correlation matrix of the observable shocks $\Sigma$, estimated before the model, and for $\Sigma_0$ to remain p.s.d. during model estimation, as it changes with different parameter guesses of $\sigma_{z\nu}$.
**Likelihood.** In our model, each i.i.d. observation is a sequence of search actions followed by a terminal action performed by a consumer. Hence, the likelihood function of interest characterizes the probability of a sequence of actions conditional on a set of parameters and data. It is given by

\[
L(\theta | X, A_i) = \prod_{i=1}^{N} Pr(A_i | \theta, X)
\]  

(21)

where \( X \) contains vehicle characteristic data and the pre-estimated correlation matrix \( \Sigma \), \( A_i = \{a_1^i, ..., a_{T_i}^i\} \) is the sequence of actions of individual \( i \), \( T_i \) is the number of actions taken by individual \( i \), and \( \theta \) is a vector of parameters. Because the shocks \( \nu_{ij} \) are unknown to the econometricians, we employ a simulated maximum likelihood approach:

\[
\tilde{Pr}(A_i | \theta) = \frac{1}{R} \sum_{r=1}^{R} Pr(A_i | \theta, \nu_i^r)
\]  

(22)

and the simulated log-likelihood for the sample follows:

\[
l(\theta | X) = \log \left( \prod_{i=1}^{N} \tilde{Pr}(A_i | \theta) \right)
\]

\[
= \sum_{i=1}^{N} \log \left( \tilde{Pr}(A_i | \theta) \right)
\]  

(23)

### 3.5 Identification

The identification of the search model is relatively complex, because most of the parameters affect each of the moments of the corresponding data generation process. The preference parameters are inferred through the relationships between their respective regressors and purchase rates, as in traditional discrete choice models. However, they also affect search behaviors. Notice, for example, that a consumer may be more willing to search a vehicle’s characteristic, say \( x_1 \), when her corresponding preference weight \( \beta_1 \) is higher. The reason is that higher values of \( \beta_1 \) induce higher variances of the expected search utility from evaluating \( x_1 \).

The search patterns in the data, together with the conversion decisions, allow us to characterize the preference parameters and the search cost parameters. The latter ones are intuitively identified by the average propensity of consumers to take each search action, conditional on the covariance parameters. This last statement is especially applicable to
search actions involving vehicle characteristics that are known to the econometricians. In
terms of the unobservable characteristic, there is an observational equivalence between a
consumer who faces a low search cost and low variance with another who faces a high search
cost and high variance surrounding the same characteristic. Simply put, more uncertainty
surrounding a characteristic increases the option value of search, which in turn can be offset
by higher search costs. In order to take this into account, we normalize the variance of the
unobservable characteristic to one.\footnote{This normalization applies to virtually all models of search for unobserved components.}

It remains to discuss the identification of the covariance parameters between the observed
characteristics and the unobserved utility component $\nu$. Whereas search costs are identified
by the average propensity of each search action being taken, the covariance parameters are
identified by information-dependent search patterns concerning the unobserved characteristic.
For example, suppose a consumer knows a given vehicle’s price. One can imagine that
if the price is moderately high, she may opt for the outside option immediately, conditional
on the preference and search cost parameters. If the price were lower, it is possible that the
consumer would have continued her search of the vehicle’s attributes, including component
$\nu$. The price threshold at which the consumer is indifferent between continuing her search
for $\nu$ identifies the covariance between price and the unobserved component: If the consumer
selected the outside option without searching further, despite facing only a moderate price
point, then it must have been that the residual variance of the unknown component was
not high enough to justify further search. On the other hand, continuing search despite the
current low prior on the vehicle’s expected utility is a signal of high variance of component
$\nu$, conditional on the known characteristics. The parameters $\sigma_{z\nu}$ of the variance-covariance
matrix thus identify the residual variance of component $\nu$, conditional on knowledge of the
vehicles’ remaining characteristics.

4 Results

Before presenting the model estimates, we detail the correlation structure among the observ-
able vehicle characteristics, net of the make/model and color effects. Table 6 presents the
cross-correlations for the residuals discussed in equation (15). The results are in line with
those in Table 2, which presents the correlations before the effects of make/model and color
are removed. The main difference is related to price: The negative correlation between price and the remaining characteristics tends to increase as a result of removing the effects of the categorical variables.

Table 7 presents the estimates of the preference coefficients. Focusing on the point estimates first, we find that consumers dislike more expensive vehicles, with higher mileage, higher number of inspection issues, older, involved in more accidents and with more previous owners. Of the preference parameter estimates, Mileage and the Number of Accidents do not appear as statistically significant.\footnote{Statistical significance of some preference parameters also disappeared in the Logit regression, after the inclusion of make-model and color indicators. The p-values associated with Mileage and Number of Accidents are 0.307 and 0.104, respectively.} Overall, the results are in line with general expectation.

Table 11 provides a comparison between the model predictions and some key moments in the data. Overall, the model matches average search intensities well, although it tends to underestimate the maximum search activities observed in the data. In practice, model fit is satisfactory, especially given that the model’s parameters are estimated via likelihood maximization rather than through matching the moments in the data directly.

Table 8 presents the estimates of the search cost parameters by segment. The size of the first segment is of approximately 89%. These consumers exhibit lowest search costs for photos, and the highest search cost is related to inspecting vehicle inspection reports. This ordering is highly consistent with the layout of the vehicle detail page at the time of the data collection, where the link to the vehicle’s inspection report was located at the bottom of the vehicle detail page, the vehicle photos were located exactly on top, with the link to the vehicle history just below. The second segment is much smaller, and exhibits a different pattern, in that inspection reports are prioritized against the other two features. A few notes are in order. First, the value of the search costs must be interpreted as a function of the normalized variance of the unknown utility component, \( \sigma^2 \nu \). As the researchers set \( \sigma^2 \nu \) to a higher value, the estimated search costs will also rise, as the model attempts to explain why consumers do not search photos as much, given the high option value implied by \( \sigma^2 \nu \). Similarly, when \( \sigma^2 \nu \) is normalized to a low number, search costs decrease, possibly becoming negative. While there exists relative information across search cost estimates, one should keep in mind that their magnitudes reflect the normalization of the variance of the unobserved search component, and so a negative estimate should not be immediately interpreted as a search ‘benefit.’
These results point to two very different segments: The first prioritizes browsing vehicle pictures, which provide more subjective information and hedonic utility, whereas relatively speaking, the second segment prefers more objective information, as provided by the vehicle’s history and the inspection report.

The estimates of the variance parameters are presented in Table 9. They then yield the estimated variance-covariance matrix presented in Table 10, which moderates the relationships across characteristics, including $\nu$. All covariances of the observable characteristics, when parsed out of the categorical variables effects', are statistically significant at the 1% level, with sensible directions. We recover low cross-correlations with the unobserved component, which is interpreted to mean that vehicle pictures mostly reveal idiosyncratic information, in terms of the relation to the observable characteristics. We find opposing significant small correlation estimates between the unobserved utility component and the observable characteristics of mileage and number of accidents.

5 Counterfactual Analyses

5.1 Designing Information Provision

Stylized Example. The effects of information provision decisions on profitability are not immediately intuitive. When consumers are strategic and understand the distribution of characteristics in the product set they face, should they not anticipate the average benefits of searching, independently of how information is organized?

We first provide an example that illustrates that the decision of which characteristics are immediately available to consumers 1) can have performance implications for outcomes and 2) is challenging to predict a priori. In this example, we consider a seller offering 4 vehicles, each with two characteristics, according to:
Table A: Example of Seller Offering Four Vehicles

<table>
<thead>
<tr>
<th></th>
<th>Vehicle A</th>
<th>Vehicle B</th>
<th>Vehicle C</th>
<th>Vehicle D</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>+20</td>
<td>+20</td>
<td>-20</td>
<td>-20</td>
</tr>
<tr>
<td>$x_2$</td>
<td>-10</td>
<td>+10</td>
<td>-10</td>
<td>+10</td>
</tr>
<tr>
<td>$u = x_1 + x_2$</td>
<td>+10</td>
<td>+30</td>
<td>-30</td>
<td>-10</td>
</tr>
</tbody>
</table>

and each vehicle may be in one consumer’s consideration set with equal probability (equivalently, there may exist four consumers, and each searches a different alternative). When consumers have perfect information, they derive conversion utility of $u = x_1 + x_2$.

The first two rows of Table A denote the vehicles’ characteristics, and rows a)-d) denote a potential buyer’s utility, at different information sets. If the consumer knows nothing, then her expected utility is equal to zero, because in this example each positive characteristic of a vehicle is exactly offset by the negative characteristic of another.\(^{23}\)

If the consumer decides to search characteristic 1, she updates her utility to -20 or 20, depending on the vehicle she is inspecting. If she searches characteristic 2, her utility is updated to -10 or 10, because characteristic 2 exhibits lower variance than characteristic 1. If the consumer searches both characteristics, then she will find herself in one of four possible distinct cases, with equal probability. Finally, we assume there exists an outside option, with known value of 15 to the consumer.

Consider how the firm’s profit depends on the information design. If the firm provides perfect information, only vehicle B is sold, yielding a conversion rate of 25%. The fact that information provision is beneficial to the firm can be easily seen by considering the case in which the firm hides both attributes, so that consumers expect an utility of zero and thus never buy.

In our example, vehicles C and D never yield expected utility above the outside option, and so are never bought irrespective of the consumer’s information set when searching them. Vehicle B is attractive as long as the consumer knows $x_1$ or both characteristics, and vehicle A is attractive as long as the consumer knows $x_1$ only. We now show that the firm may

\(^{23}\)In reality, the consumer’s idealized distribution may be different from the realized distribution of characteristics offered by the seller. Our example is stricter, in that the consumer’s beliefs equal the seller’s offering exactly.
be able to induce purchases of both of these vehicles, by selecting which characteristics to
feature.

Suppose the seller decides to feature characteristic $x_1$, but not $x_2$. In this case, both
vehicles A and B (at least momentarily) yield a high enough utility to be bought. Then, the
consumer should search further if and only if the search cost is low enough:

$$-c + Pr(x_2 = -10) \cdot 15 + Pr(x_2 = 10) \cdot 15 \geq 20$$

$$\iff c \leq 2.5$$

When $c > 2.5$, the consumer stops her search after learning $x_1$, and the seller’s conversion
rate increases to 50%. The gain of the seller arises because it induces search so that some
consumers find positive aspects, but then precludes search, so that consumers do not realize
their utility fully. The seller benefits by ensuring that a sliver of consumers is willing to
convert before acquiring full information.

The opposite strategy, in which the seller only discloses information about characteristic
$x_2$, may not increase profits when search costs are low enough. Upon knowing $x_2$, consumers
search $x_1$ if and only if

$$-c + Pr(x_1 = -20) \cdot 15 + Pr(x_1 = 20) \cdot 30 \geq 15$$

$$\iff c \leq 7.5$$

In this case, there exist higher levels of search costs such that consumers prefer to search
further. Once they do, they will obtain perfect information, and the seller’s conversion rate
will once again equal 25%.

In summary, the seller may have an incentive to first disclose the characteristic with the
highest variance, in order to induce some consumers to hold an interim high utility, and not
search further. These consumers will convert despite the possibility of earning a lower utility
than their outside option, ex post. Table B summarizes the seller’s payoffs for the different
levels of search costs.
Table B: Conversion Rates, as a Function of Information Disclosure Strategy and Search Costs

<table>
<thead>
<tr>
<th></th>
<th>$c \leq 2.5$</th>
<th>$c \in (2.5, 7.5]$</th>
<th>$c &gt; 7.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Information</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Perfect Information</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>Disclose $x_1$; $x_2$ is search characteristic</td>
<td>0.25</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Disclose $x_2$; $x_1$ is search characteristic</td>
<td>0.25</td>
<td>0.25</td>
<td>0</td>
</tr>
</tbody>
</table>

Table B reveals intricate effects of the information provision strategy on firm outcomes. Firm profits can increase with search costs when only characteristic $x_1$ is revealed (the high variance characteristic) and can decrease with search costs when only characteristic $x_2$ is revealed (the low variance characteristic). Offering some information may bring conversion rates up (compare the first row of Table B with any other): by providing enough information, some consumers will want to remain in the platform and explore further. However, too much information (second row of Table B) may partition consumer utility too much, providing the chance to find all the negatives along with the positives.

The example above illustrates that the optimal information design is challenging to decide in practice, based on simple heuristics or observations, especially once one considers additional aspects such as more complicated joint distributions of product characteristics, preference parameters or different search costs. This in turn highlights the need of considering formal counterfactual analyses in order to assess the effects of different information design strategies.

**Designing Information Provision.** We recompute consumers’ search decisions and conversion actions while varying the information structure. We first exchange characteristics listed on the main page with those in the vehicle detail page, one by one. These analyses are motivated by the possibility that the firm may not be able to maintain consumers’ cognitive loads relatively constant if it introduces too much information in one place. For example, moving all characteristic information to the main listing page could potentially make the firm’s website liken a spreadsheet, severely damaging the user’s experience. In the same vein, we do not consider the counterfactual scenario of moving a vehicle’s photo gallery to the main listing page.

---

24 We focus the analysis on the seller, because the buyer is always better off with corner case of perfect information.
Table 12 presents the effects of exchanging characteristics between the main listing page and the vehicle detail page, one by one, on conversion rates. The fact that in the original seller’s website the vehicle history information is located up on the page, while the inspection report is located near the bottom, suggests that there may be different effects of exchanging each of the ex ante characteristics with each of the search characteristics (and their associated positions). Hence, we inspect all of the one-to-one exchanges.

Focusing first on the first two rows of Table 12, we find statistically significant conversion effects, related to making vehicles’ ages search characteristics. The relatively modest values are consistent with the interpretation that information design can affect profitability by influencing consumers who are close to being indifferent between searching further. It is also in line with the relatively expensive item values, when compared with the search costs of browsing a seller’s website.

Exchanging vehicle mileage with a search characteristic has estimated negative conversion effects, whereas exchanging vehicle age with the same characteristics can increase conversion rates by up to 1.65%. The latter changes are statistically significant, such that there is scope to improve outcomes through the design of information provision, even in the context of big ticket items like used cars.

In the third row of Table 12, we report the effects of moving pricing information to the vehicle detail page. We include this counterfactual scenario not because of its effects, but because it illustrates that some counterfactuals may fall outside the scope of the model. Specifically, shrouding price information by relegating it to a search characteristic may introduce equilibrium effects that fall outside scope of our model: Consumers could rationalize such a decision by assuming that such a seller cannot be trusted, or is attempting to overprice vehicles by disclosing pricing information as late as possible. This scenario is valuable in that the model predicts that seller outcomes can improve by reducing the amount of information provided to consumers initially (price is the most informative attribute).

To investigate this issue further, we report the effects of the different information scenarios on search in Table 13. We find that the counterfactual policies decrease consumers’ search behaviors. When compared with the conversion rate implications presented in Table 13, our results highlight the fact that the effects of disclosing information are not clearly predictable ex ante, and cannot be predicted by the resulting search intensities. This result is reinforced by the results of the analysis under full information, which we discuss later in this section.
We report the exhaustive list of counterfactual scenarios, beyond the design decision of exchanging one attribute in the main listing page with one attribute in the vehicle detail page, in Table 14. For example, scenario a) in Table 14 considers scenario

\[
\begin{bmatrix}
\text{Mileage} \\
\text{Age}
\end{bmatrix}
\text{Ex-Ante};
\begin{bmatrix}
\text{V. Hist.} \\
\text{Insp. Rep.}
\end{bmatrix}
\text{VehicleDetailPage}
\rightarrow
\begin{bmatrix}
\text{V. Hist.} \\
\text{Insp. Rep.}
\end{bmatrix}
\text{Ex-Ante};
\begin{bmatrix}
\text{Mileage} \\
\text{Age}
\end{bmatrix}
\text{VehicleDetailPage}
\]

which examines the conversion rate implications of moving both the vehicle history and inspection notes information to the main listing page, and replacing vehicle history information with mileage information, and the inspection report with vehicle age information. Table 14 reveals that the most profitable information configuration is Scenario e), in which vehicles’ ages are moved to the vehicle detail page, the vehicle’s inspection report is brought to the main listing page, and the vehicle history is displayed below vehicle age in the vehicle detail page. The conversion impact is of +0.08%, a 1.55% relative increase.

To conclude, we consider the scenario where search costs are very low, which yields the perfect information case. While it is admittedly an idealized scenario, it is nonetheless an informative benchmark. We present the results in Table 15. The perfect information scenario is estimated to increase the conversion rate from 4.9% to 5.4%, the largest increase we have observed across counterfactual analyses. The results suggest that seller outcomes are constrained by the amount of information it is able to provide costlessly to consumers. Unlike in the main counterfactual analysis, which predicted statistically significant gains by inducing lower search activity, the perfect information scenario induces full search and higher conversion rates. This result also highlights the importance of conducting case-by-case analyses in order to assess the effect of information design on conversion rates, since search activity is not a perfect predictor of firm performance. Rather, consistent with our illustrative example from before, the analysis shows that changing the information design in order to restrict or encourage consumer search does not produce direct predictions, positive or negative, on conversion rates.
5.2 Knowledge Gradient

Our dynamic problem can be adapted to produce knowledge gradient estimates. The knowledge gradient method assumes the decision-maker acts as if she behaved much like a standard numerical optimizer, seeking the action of highest immediate value compared to the search costs. Under the knowledge gradient assumption, terminal utilities are calculated as before:

\[ V_j(\Omega_i, \epsilon_i) = E(v_{ij}|\Omega_i, \epsilon_i) \]  

(25)

but the dynamic utility from search decisions takes on a different form:

\[ V_{j,s}(\Omega_i, \epsilon_i) = -c_s + E_{\epsilon_i, \omega js}\left[ \max \left\{ v_{i0}, V_1(\Omega_i, \epsilon_i'), \ldots, V_j(\Omega_i \cup \omega js, \epsilon_i'), \ldots, V_J(\Omega_i, \epsilon_i') \right\} \right] \Omega_i + \epsilon_{ij} \]

(26)

The formula above reveals that decision makers following the knowledge gradient rule behave as if each search was their last one, and as if they would be forced to make a terminal decision immediately afterwards. The knowledge gradient literature takes advantage of the ‘steepest ascent’ idea, in terms of the immediate option value of searching relative to the search cost.

We adapt our value functions by recalculating search values according to (26), and re-estimate the model. We compare the predicted moments of the knowledge gradient model with the ones in the data in Table 16. The fit measures can be compared with those of the full forward-looking case, presented in Table 11. Across the board, the knowledge gradient approach fits the sample moments worse than the forward-looking model. This occurs despite the fact that the models share the same number of parameters.

We conduct a non-nested test à la Vuong (1989), in order to verify whether the knowledge gradient and the forward-looking models are at the same ‘statistical distance’ from some true underlying data generation process. The test rejects that the models are at the same distance from the true data generation process with p-value less than 1% (see Table 17).

Overall, our results suggest that, despite its convenience and lighter estimation burden, the knowledge gradient approach appears to provide a poor approximation to the forward-looking case. In contrast with some theoretical results and other empirical applications (see Introduction section for relevant work), our results highlight the need for researchers to consider the full continuation values of search activities. In our case, failing to do so results
in worse fit and in a statistically-significant imperfect approximation to the full dynamic problem.

6 Conclusion

We have developed and estimated a model in which consumers search across attributes of multiple alternatives before making their conversion decisions. The model is able to characterize complex search patterns, including piecemeal search within each alternative and alternative revisiting. In addition, it allows for a rich joint distribution of the observable characteristics, as well as the characterization of the cross-correlations with an unobserved utility component. We recover the joint distribution of characteristics through a flexible approach, and use it to inform consumer beliefs. In this way, we capture the idea that good news about a specific attribute has implications for the remaining ones in consumers’ minds.

Another contribution of this paper is to consider the problem of designing information provision to consumers. By recomputing counterfactual search paths depending on different information sets, our model predicts that there are statistically significant impacts on conversion rates from introducing changes to the information environment. We find that information design decisions can affect firm outcomes significantly, the effects ranging from -0.39% to +1.65% in terms of relative changes in conversion rates. If keeping all else fixed (i.e., cognitive load, etc.), then making all information available ex ante to consumers increases the conversion rate from a prediction of 4.9% to 5.7%. Both our illustrative example as well as our results suggest that the underlying forces driving conversions are complex, and simple indicators like search intensities are poor predictors of the effect of different information designs on seller outcomes. We also compare our findings with the ones predicted by the knowledge gradient literature. Taking the dynamics of the search problem into account improves fit in terms of a number of moments in the data, and the likelihoods of the models are statistically different, implying that the myopic model provides an imperfect approximation to the full forward-looking problem.
# Figures

Figure 1: Screenshot of Main Listing Page, shift.com, October 2016
If Lucie Fox had designed the McLaren MP4-12C, the only thing he would have changed about this Wayne-worthy supercar is its slightly clumsy moniker. Everything else, from its smoldering exterior hue and singular styling to its unbelievable power, would be well-suited to the demands of a daytime Bond villain and night-time vigilante. This McLaren, with 522 horses, seven speeds, and a smattering of thoughtfully placed knobs and buttons to its name, is a childhood fantasy sprung to life. It lets black-clad do-gooders like you zoom around in stealthy, sporty splendor. Its Proactive Chassis Control suspension allows bespectacled Bruce to “normal” mode and disguised supermen in “track” mode. You can enjoy the comfort of an “everyday” drive (kind of), alongside the thrill of absolutely mind-blowing performance numbers (try 0-60 in less than three seconds). In other words, this fantastical coupe delivers a spectrum of performance versatility, with Bluetooth connectivity, a backup camera, and heated seats to keep you and your sidekick comfortable and grounded. Stop ignoring that giant M projected in the night sky and follow your destiny to the wheel of this 2012 McLaren MP4-12C.
Figure 3: Densities of Avg. Session Times Across Users

Kernel Densities of Avg. Session Time
Red= Did not convert; Blue= Converted

Note: Density upper bounds selected to ensure columns are visually distinguishable across conditions.
Figure 4: Histograms and Cross Scatter Plots of Vehicle Characteristics

Note: Above, histograms in main diagonal and scatter plots in off-diagonals. Prices in (10^{-4} USD) and mileage in (10^{-5} miles).
Figure 5: Ordered Probit Regions Associated with Levels of Inspection Issues
Figure 6: Example of a State-Partitioning Decision Tree

<table>
<thead>
<tr>
<th>#</th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>
Figure 7: Example of Decision Tree

Note: Decision tree, generated for consumer searching for one vehicle, and estimation with 50 simulations.
### Tables

#### Table 1: Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vehicle Detail Pages Viewed</td>
<td>2.28</td>
<td>3.47</td>
<td>1</td>
<td>129</td>
</tr>
<tr>
<td>Number of Sessions</td>
<td>6.03</td>
<td>21.14</td>
<td>1</td>
<td>762</td>
</tr>
<tr>
<td>Photo Sets Browsed</td>
<td>1.63</td>
<td>2.65</td>
<td>0</td>
<td>74</td>
</tr>
<tr>
<td>Vehicle Histories Browsed</td>
<td>0.69</td>
<td>2.09</td>
<td>0</td>
<td>101</td>
</tr>
<tr>
<td>Inspection Reports Browsed</td>
<td>0.30</td>
<td>1.27</td>
<td>0</td>
<td>53</td>
</tr>
<tr>
<td>Average Session Duration (min.)</td>
<td>14.2</td>
<td>13.39</td>
<td>0.1</td>
<td>282.04</td>
</tr>
<tr>
<td>Activity Range (days)</td>
<td>18.67</td>
<td>38.5</td>
<td>0</td>
<td>232</td>
</tr>
</tbody>
</table>

Number of users: 24,116

Note: Statistics are per user. The event “Photo Sets Browsed” equals one for each user-vehicle observation whenever a user browsed more 20 or more pictures of a vehicle.

#### Table 2: Correlation Table for Non-Categorical Characteristics

<table>
<thead>
<tr>
<th></th>
<th>Price</th>
<th>Mileage</th>
<th>Insp. notes</th>
<th>Vehicle age</th>
<th>No. accidents</th>
<th>No. owners</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mileage</td>
<td>-0.45</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insp. notes</td>
<td>-0.16</td>
<td>0.39</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vehicle age</td>
<td>-0.40</td>
<td>0.71</td>
<td>0.44</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. accidents</td>
<td>-0.07</td>
<td>0.10</td>
<td>0.09</td>
<td>0.12</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>No. owners</td>
<td>-0.08</td>
<td>0.21</td>
<td>0.09</td>
<td>0.17</td>
<td>0.09</td>
<td>1</td>
</tr>
</tbody>
</table>

Note: Number of vehicles: 1,573. All correlation estimates statistically significant, with p-values below 0.01.
Table 3: Logit Preference Coefficient Estimates

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Point Estimates</th>
<th>Odds Ratios</th>
</tr>
</thead>
<tbody>
<tr>
<td>price</td>
<td>-1.359**</td>
<td>0.257**</td>
</tr>
<tr>
<td></td>
<td>(0.035)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>mileage</td>
<td>-1.189**</td>
<td>0.305**</td>
</tr>
<tr>
<td></td>
<td>(0.136)</td>
<td>(0.041)</td>
</tr>
<tr>
<td>age</td>
<td>-0.125**</td>
<td>0.883**</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.012)</td>
</tr>
<tr>
<td>inspection notes</td>
<td>-0.06*</td>
<td>0.942*</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.025)</td>
</tr>
<tr>
<td>n_accidents</td>
<td>-0.167†</td>
<td>0.846†</td>
</tr>
<tr>
<td></td>
<td>(0.093)</td>
<td>(0.079)</td>
</tr>
<tr>
<td>n_owners</td>
<td>-0.148**</td>
<td>0.863**</td>
</tr>
<tr>
<td></td>
<td>(0.029)</td>
<td>(0.025)</td>
</tr>
</tbody>
</table>

Number of users: 24,166

Standard errors shown in parentheses. Significance levels: † p≤0.10, * p≤0.05, ** p≤0.01. Price in (10^{-4} USD) and mileage in (10^{-6} miles).

Table 4: Vehicle Information and State Variables

<table>
<thead>
<tr>
<th>Knowledge Level</th>
<th>Characteristic</th>
<th>Utility Component</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ex-ante Characteristics</strong></td>
<td>Make/Model</td>
<td>make_model_{j}</td>
</tr>
<tr>
<td>Observable ex-ante</td>
<td>Price</td>
<td>price_{j}</td>
</tr>
<tr>
<td>&gt;&gt;</td>
<td>Mileage</td>
<td>mileage_{j}</td>
</tr>
<tr>
<td>&gt;&gt;</td>
<td>Age</td>
<td>age_{j}</td>
</tr>
<tr>
<td>&gt;&gt;</td>
<td>Color</td>
<td>color_{j}</td>
</tr>
</tbody>
</table>

**Perceived Quality/Fit**

Learned after browsing one photo set

Vehicle Photos

\( \nu_{ij} \)

**Vehicle History**

Learned upon visiting a vehicle’s history

Number of Previous Owners

owners_{j}

Number of Accidents

accidents_{j}

**Vehicle Status**

Learned upon visiting inspection reports

Number of Inspection Notes

notes_{j}

Notes: Vehicle indexes omitted. Ex-ante characteristics belong to users’ initial information set. The remaining characteristics are revealed through search actions.
Table 5: Correspondence of Search Actions and Learned Characteristics

<table>
<thead>
<tr>
<th>Search Action (s)</th>
<th>Meaning</th>
<th>Learned Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Vehicle Photos</td>
<td>$\nu_{ij}$</td>
</tr>
<tr>
<td>2</td>
<td>Vehicle History</td>
<td>$owners_j, acc_j$</td>
</tr>
<tr>
<td>3</td>
<td>Inspection Report</td>
<td>$notes_j$</td>
</tr>
</tbody>
</table>

Table 6: Estimated Correlation Table for the Errors Orthogonal to the Categorical Characteristics

<table>
<thead>
<tr>
<th></th>
<th>Price</th>
<th>Mileage</th>
<th>Insp. notes</th>
<th>Vehicle age</th>
<th>No. accidents</th>
<th>No. owners</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mileage</td>
<td>-0.66</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insp. notes</td>
<td>-0.31</td>
<td>0.37</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vehicle age</td>
<td>-0.73</td>
<td>0.71</td>
<td>0.41</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. accidents</td>
<td>-0.07</td>
<td>0.08</td>
<td>0.06</td>
<td>0.09</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>No. owners</td>
<td>-0.16</td>
<td>0.21</td>
<td>0.08</td>
<td>0.17</td>
<td>0.03</td>
<td>1</td>
</tr>
</tbody>
</table>

Note: Number of vehicles: 1,573. All correlation estimates significant with p-values below 0.01.

Table 7: Model Estimates: Preference Parameters

<table>
<thead>
<tr>
<th></th>
<th>Parameter Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vehicle Characteristics</strong></td>
<td></td>
</tr>
<tr>
<td><em>price</em></td>
<td>-0.197** (0.017)</td>
</tr>
<tr>
<td><em>mileage</em></td>
<td>-0.099 (0.097)</td>
</tr>
<tr>
<td><em>notes</em></td>
<td>-0.148* (0.063)</td>
</tr>
<tr>
<td><em>age</em></td>
<td>-0.049** (0.012)</td>
</tr>
<tr>
<td><em>accidents</em></td>
<td>-0.318 (0.196)</td>
</tr>
<tr>
<td><em>owners</em></td>
<td>-0.188** (0.071)</td>
</tr>
</tbody>
</table>

Log-Likelihood: -40,421.99

Make-Model and Color Dummies ✓

N= 12,887

Note: Standard errors in parentheses. Significance levels:
$\dagger p\leq0.10$, $^* p\leq0.05$, $^{**} p\leq0.01$. 

49
Table 8: Model Estimates: Search Cost Parameters

<table>
<thead>
<tr>
<th>Parameter Estimates</th>
<th>Segment 1</th>
<th>Segment 2</th>
<th>Other Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$c_{photos}$</td>
<td></td>
<td>$\alpha$</td>
</tr>
<tr>
<td></td>
<td>0.05**</td>
<td>2.552**</td>
<td>-2.136**</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.054)</td>
<td>(0.032)</td>
</tr>
<tr>
<td></td>
<td>$c_{vehicle, history}$</td>
<td>2.259**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.127)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$c_{insp., report}$</td>
<td>5.534**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.045)</td>
<td>(0.144)</td>
<td></td>
</tr>
<tr>
<td>Other Parameters</td>
<td>Make-Model and Color Dummies</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N= 12,887</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Standard errors in parentheses. Significance levels: † $p \leq 0.10$, * $p \leq 0.05$, ** $p \leq 0.01$. Size of segment 1, implied by $\hat{\alpha}$, is 89.43%.
Table 9: Model Estimates: Variance-Covariance Parameters

<table>
<thead>
<tr>
<th>Parameter Estimates</th>
<th>Variance-Covariance Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.002</td>
<td>s1</td>
</tr>
<tr>
<td>(0.061)</td>
<td>s2</td>
</tr>
<tr>
<td>0.079**</td>
<td>s3</td>
</tr>
<tr>
<td>(0.001)</td>
<td>s4</td>
</tr>
<tr>
<td>0.053</td>
<td>s5</td>
</tr>
<tr>
<td>(0.131)</td>
<td>s6</td>
</tr>
<tr>
<td>0.022</td>
<td></td>
</tr>
<tr>
<td>(0.061)</td>
<td></td>
</tr>
<tr>
<td>-0.081**</td>
<td></td>
</tr>
<tr>
<td>(0.026)</td>
<td></td>
</tr>
<tr>
<td>0.025</td>
<td></td>
</tr>
<tr>
<td>(0.069)</td>
<td></td>
</tr>
</tbody>
</table>

Make-Model and Color Dummies ✓
N= 12,887
Note: Standard errors in parentheses. Significance levels: † p≤0.10, * p≤0.05, ** p≤0.01.

Table 10: Estimated Variance-Covariance Matrix

\[
\Sigma_0 = \begin{pmatrix}
1 & -0.66 & -0.31 & -0.73 & -0.07 & -0.21 & 0.00 \\
-0.66 & 1 & 0.37 & 0.71 & 0.08 & 0.27 & 0.08** \\
-0.31 & 0.37 & 1 & 0.41 & 0.07 & 0.14 & 0.05 \\
-0.73 & 0.71 & 0.41 & 1 & 0.09 & 0.24 & 0.02 \\
-0.07 & 0.08 & 0.07 & 0.09 & 1 & 0.021 & -0.08** \\
-0.21 & 0.27 & 0.14 & 0.24 & 0.021 & 1 & 0.03 \\
0.00 & 0.08** & 0.05 & 0.02 & -0.08** & 0.03 & 1
\end{pmatrix}
\]

<table>
<thead>
<tr>
<th>Price</th>
<th>Mileage</th>
<th>Insp.notes</th>
<th>Vehicle age</th>
<th>No.accidents</th>
<th>No.owners</th>
<th>ν</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.08**</td>
<td>0.05</td>
<td>0.02</td>
<td>-0.08**</td>
<td>0.03</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Note: In bold, cross-correlation elements induced by estimated parameters s1..s6. Rightmost vector shows the corresponding characteristics. Non-bold estimates all significant at 1% level. Coefficients in bold: † p≤0.10, * p≤0.05, ** p≤0.01.

51
Table 11: Search Moments from Dataset and Model Predictions

<table>
<thead>
<tr>
<th></th>
<th>Data</th>
<th>Model Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Dev.</td>
</tr>
<tr>
<td>Conversion Rate</td>
<td>0.047</td>
<td>0.213</td>
</tr>
<tr>
<td>Avg. N. Searches</td>
<td>1.639</td>
<td>3.22</td>
</tr>
<tr>
<td>Vehicle Histories</td>
<td>0.441</td>
<td>0.743</td>
</tr>
<tr>
<td>Inspection Reports</td>
<td>0.175</td>
<td>0.496</td>
</tr>
<tr>
<td>Photo Sets</td>
<td>1.023</td>
<td>0.873</td>
</tr>
</tbody>
</table>

N: 12,887 consumers

Note: All model predictions are first averaged across simulations, and then statistics are applied. The Standard Deviation, Min and Max statistics of the Conversion Rate were corrected through use of the Bernoulli distribution moments (e.g., Std. Dev. = $\sqrt{p(1-p)}$).

Table 12: Conversion Effects of Exchanging Attributes Between Front and Detail Pages

<table>
<thead>
<tr>
<th>Vehicle Detail Page Attribute</th>
<th>Front Page Attribute</th>
<th>Vehicle History</th>
<th>Inspection Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mileage</td>
<td>-0.01% (-0.26%)</td>
<td>-0.02% (-0.39%)</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>+0.06% (+1.28%)**</td>
<td>+0.08% (+1.65%)**</td>
<td></td>
</tr>
<tr>
<td>Price (see note)</td>
<td>+0.07% (+1.38%)**</td>
<td>+0.07% (+1.35%)**</td>
<td></td>
</tr>
</tbody>
</table>

N: 12,887 consumers

Note: Above, impact of exchanging attributes in the front page with the ones in the vehicle detail page on mean conversions. Significance levels: † p≤0.10, * p≤0.05, ** p≤0.01. The scenario related to price is introduced to illustrate that not all counterfactual analyses are necessarily addressable by our model.

Table 13: Search Effects of Exchanging Attributes Between Front and Detail Pages

<table>
<thead>
<tr>
<th>Vehicle Detail Page Attribute</th>
<th>Front Page Attribute</th>
<th>Vehicle History</th>
<th>Inspection Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mileage</td>
<td>-0.03% (-0.02%)†</td>
<td>-0.01% (-0.00%)</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>-0.03% (-0.02%)†</td>
<td>0.00% (0.00%)</td>
<td></td>
</tr>
<tr>
<td>Price</td>
<td>-0.02% (-0.01%)†</td>
<td>-0.01% (-0.01%)</td>
<td></td>
</tr>
</tbody>
</table>

N: 12,887 consumers

Note: Above, impact of exchanging attributes in the front page with the ones in the vehicle detail page on the number of searches. Significance levels: † p≤0.10, * p≤0.05, ** p≤0.01.
Table 14: Conversion Effects of Different Information Configurations

<table>
<thead>
<tr>
<th>Information Configuration</th>
<th>Conversion Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original configuration:</td>
<td></td>
</tr>
<tr>
<td>Main Listing Page</td>
<td></td>
</tr>
<tr>
<td>[Mileage]</td>
<td></td>
</tr>
<tr>
<td>[Age]</td>
<td></td>
</tr>
<tr>
<td>Vehicle Detail Page</td>
<td></td>
</tr>
<tr>
<td>[V. Hist.]</td>
<td></td>
</tr>
<tr>
<td>[Insp. Rep.]</td>
<td></td>
</tr>
</tbody>
</table>

Scenario a)  
[ V. Hist. ]  
[ Insp. Rep. ]  
[Mileage]  
[Age]  
+0.01% (+0.13%)

Scenario b)  
[ V. Hist. ]  
[ Insp. Rep. ]  
[Age]  
[Mileage]  
-0.01% (-0.21%)

Scenario c)  
[Mileage]  
[Age]  
[Insp. Rep.]  
[V. Hist.]  
+0.02% (+0.47%)

Scenario d)  
[Insp. Rep.]  
[Age]  
[Mileage]  
[V. Hist.]  
-0.00% (-0.01%)

Scenario e)  
[Mileage]  
[Insp. Rep.]  
[Age]  
[V. Hist.]  
+0.08% (1.55%)**

Scenario f)  
[V. Hist]  
[Age]  
[Insp. Rep.]  
[Mileage]  
-0.01% (-0.26%)

Scenario g)  
[Mileage]  
[V. Hist]  
[Insp. Rep.]  
[Age]  
+0.06% (+1.27%)**

N: 12,887 consumers

Note: Above, impact of exchanging attributes in the front page with the ones in the vehicle detail page on the number of searches. Significance levels: † p≤0.10, * p≤0.05, ** p≤0.01.
### Table 15: Consumer Behavior Statistics Compared with Perfect Information Benchmark

<table>
<thead>
<tr>
<th></th>
<th>With Search Costs</th>
<th>Perfect Information</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Dev.</td>
</tr>
<tr>
<td>Conversion Rate</td>
<td>0.049</td>
<td>0.216</td>
</tr>
<tr>
<td>Avg. N. Searches</td>
<td>1.639</td>
<td>1.286</td>
</tr>
<tr>
<td>Vehicle Histories</td>
<td>0.436</td>
<td>0.437</td>
</tr>
<tr>
<td>Inspection Reports</td>
<td>0.177</td>
<td>0.123</td>
</tr>
<tr>
<td>Photo Sets</td>
<td>1.027</td>
<td>0.739</td>
</tr>
</tbody>
</table>

N: 12,887 consumers

Note: All predictions are first averaged across simulations; then statistics are applied. The Standard Deviation, Min and Max statistics of the Conversion Rate were corrected through use of the Bernoulli distribution moments (e.g., Std. Dev = $\sqrt{p(1-p)}$). The search statistics in the perfect information case are bounded above by the search sets of each consumer in the dataset.

### Table 16: Consumer Behavior Statistics Compared with Knowledge Gradient

<table>
<thead>
<tr>
<th></th>
<th>Data</th>
<th>Knowledge Gradient</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Dev.</td>
</tr>
<tr>
<td>Conversion Rate</td>
<td>0.047</td>
<td>0.213</td>
</tr>
<tr>
<td>Avg. N. Searches</td>
<td>1.639</td>
<td>3.22</td>
</tr>
<tr>
<td>Vehicle Histories</td>
<td>0.441</td>
<td>0.743</td>
</tr>
<tr>
<td>Inspection Reports</td>
<td>0.175</td>
<td>0.496</td>
</tr>
<tr>
<td>Photo Sets</td>
<td>1.023</td>
<td>0.873</td>
</tr>
</tbody>
</table>

N: 12,887 consumers

Note: All predictions are first averaged across simulations; then statistics are applied. The Standard Deviation, Min and Max statistics of the Conversion Rate were corrected through use of the Bernoulli distribution moments (e.g., Std. Dev = $\sqrt{p(1-p)}$).

### Table 17: Vuong’s Non-Nested Test

<table>
<thead>
<tr>
<th></th>
<th>Log-likelihood</th>
<th>Vuong’s Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward-Looking Model</td>
<td>-40,421.99</td>
<td>$Z$-statistic $= \frac{2.116.28}{27.29} = 77.55$</td>
</tr>
<tr>
<td>Knowledge Gradient</td>
<td>-42,538.28</td>
<td>p-value $= 0.000$</td>
</tr>
</tbody>
</table>

N: 12,887
Appendix

A  Data Cleaning

A few unusual patterns emerged while analyzing the clickstream data. First, a few sessions were very short, resembling ‘bouncing’ behavior in which a user visits the website and leaves without triggering additional action flags, often adding up to a very short overall session time. This type of activity is typically generated by users who visited the website by mistake, or by ‘bots’ scraping the Internet for content. These users are likely to have very different objectives than purchasing a car on the platform. We control for this type of behavior by eliminating sessions that triggered only a page visit or that took less than 5 seconds in terms of the overall measured activity (e.g., a session that triggers two events, with only 3 seconds apart, is eliminated). We also removed a few users with very unusual browsing behaviors. For example, users who exhibited activity across multiple IP addresses belonging to different countries within a very short period of time were eliminated from the sample.

B  Variance-Covariance Matrix Decomposition

The variance-covariance matrix

$$\Sigma_0 = \begin{pmatrix}
\Sigma \\
\sigma'_{z
u} \\
\sigma_{z
u} \\
\sigma'_{\nu} \\
\sigma^2_{\nu}
\end{pmatrix}$$

has a known component $\Sigma$, given in Table 6, and an unknown border, to be estimated. We take the Cholesky decomposition of $\Sigma_0$, and parameterize its border with vector

$$\begin{bmatrix}
s_1, s_2, ..., s_6, \\
\sqrt{1 - \sum_{j=1}^{6} s_j^2}
\end{bmatrix}$$

This yields the decomposed matrix in Table 18.
Table 18: Cholesky Decomposition of Bordered Variance-Covariance Matrix

\[
C = \begin{pmatrix}
1. & -0.664415 & -0.305486 & -0.727454 & -0.0660357 & -0.159545 \\
0 & 0.747364 & 0.217517 & 0.29733 & 0.0471439 & 0.143938 \\
0 & 0 & 0.927019 & 0.128614 & 0.0370378 & 0.0018745 \\
0 & 0 & 0 & 0.604867 & 0.0352993 & 0.016755 \\
0 & 0 & 0 & 0 & 0.995389 & 0.00822777 \\
0 & 0 & 0 & 0 & 0 & 0.976461 & \sqrt{1 - \sum_{j=1}^{6} s_j^2}
\end{pmatrix}
\]

The matrix \( \Sigma \), obtained by

\[
\Sigma = C'C
\]

conforms to matrix \( \Sigma_0 \) in that the upper-left block is equal to \( \Sigma \), and the element

\[
\sqrt{1 - \sum_{j=1}^{6} s_j^2}
\]

in \( C \) implies that resulting lower-right parameter \( \sigma_\nu^2 \) is normalized to one. In addition, vector \( \sigma_{z\nu} \) in \( \Sigma_0 \) is obtained by linear combinations of parameters \( \{s_1..s_6\} \). Finally, at each iteration we check that matrix \( \Sigma \) is p.s.d., and penalize likelihood evaluations for guesses generating non-p.s.d. covariance matrices.

C  State-Partitioning Decision Trees

C.1  Creating the Decision Tree

The decision tree has three fundamental objects. The first one is a node, which corresponds to each possible “action state” (set of search actions taken) that the consumer may face during her search. By action state, we mean a value that identifies which search actions the consumer has already taken.

The second fundamental object is a partition. This object identifies a set of eligible simulations at a given node. Each node contains a set of eligible partitions, each of which the consumer may encounter at that node. For example, when the consumer searches an action with a binary value, she partitions the set of eligible simulations into two subsets, and selects one depending on the observed value. The corresponding child node features both of those partitions.

Finally, the third object is an (search) action, which belongs to a specific partition. Moreover, each partition may contain multiple admissible search actions. The action object has
two roles. First, it links a partition to its child partitions and their probabilities. For example, if a consumer takes a given search action while in a given partition, she will transition to different partitions of a sub-node, with different probabilities. Second, an action also contains its own value, which is obtained by integrating the value of the possible future paths it can lead to.

The connections among all of these objects describe the DM’s dynamic problem. Below, we provide a stylized representation of each of the fundamental decision-tree objects. (The actual representation is more complex in that involves vectors of pointers and additional properties and methods.)

**Node:**

```plaintext
node{
    int id;
    vector of <partition> v_p;
}
```

**Partition:**

```plaintext
partition{
    vector of <action> v_a;
    vector of <int> simulations;
    vector of <terminal_utilities> v_tu;
}
```

**Action:**

```plaintext
action{
    vector of <partition> v_p;
    vector of <transition_probabilities> v_tp;
    double EV;
}
```
In order to understand the links among the objects above, consider Figure 6. When the consumer starts her search, she is in partition 1 of node 1. This partition is saved as an element of the vector “v_p” of the node. The partition is associated with an object that has a vector of actions “v_a”, a vector of identifiers of the simulations it contains (field “simulations”) and a vector of terminal utilities. In the example, vector “v_tu” has two elements, corresponding to the inside and outside options. Finally, each action in vector “v_a” has three fields. First, an action keeps track of the partitions it can transition the decision-maker to (field “v_p”). It also keeps track of the transition probabilities to each of those partitions (field “v_tp”), and finally, it keeps track of its own value (field “EV”).

Consider again the tree represented in Figure 6, which has four nodes. Notice that partition $P_1$ shows up as a field of node 01, as well as a field of action 01 of node 00. Relying on pointers is helpful in not having to allocate partition $P_1$ twice. Rather, object $P_1$ is created only once, with node 00 and action 01 merely keeping references to it.

We create the decision tree bottom-up, starting at the last node corresponding to all search actions taken (in the working example, node 11). We populate the last node with single-element partitions. We then proceed to populate all direct parent nodes (in the working example, nodes 01 and 10). For each one, we cycle through the elements in the partitions of the child node (node 11), and attempt to merge them so that each partition of the parent node only differs on the characteristics that have not been searched yet. For each partition, we then create the actions, which link to the child partitions. For example, the only admissible search action allowed at node 01 is action 10. Hence, we add action 10 to each partition of node 01, and link it to the potential child partitions, taking into account the characteristics that the action reveals. The procedure stops once the top of the tree is reached.

### C.2 Updating the Decision Tree

Consider a set of structural parameters being attempted by the optimizer. The tree is updated in four stages:

1. For each consumer, given the new guess for the structural parameters, the algorithm re-draws values of $\nu$, conditional on the observed characteristics. If these values remain the same as the previous draws, a previously available decision tree can be used for the new iteration. Otherwise, a new tree needs to be generated for consumer $i$. 

58
2. Terminal action utilities are calculated for all partitions. For example, at node 00, calculating the expected value of the inside option involves performing a weighted average of the characteristics of all simulations, and performing the inner product of the resulting vector with the vector of current preference parameters.

3. The transition probabilities are recalculated for all actions. These are calculated from the simulation frequency of the simulated draws.

4. Proceeding bottom-up, the expected value of each search action is calculated, by cycling through each of the potential child partitions (the partitions that the action may lead to). For each child partition, the log-sum of all of the action utilities (inside, outside and search actions) is calculated. Finally, those expected utilities are weighted by the transition probabilities.

This process is accelerated by noticing that some steps can be skipped when the simulations remain constant across evaluations. Namely, the tree structure remains valid (it has the same partitions and actions) as long as the draws of $\nu$ are the same at the new structural parameters (e.g., when only preference parameters change across iterations). A tree from a previous evaluation can be used in this case, eliminating the need to create a new one. Moreover, noticing that the transition probabilities also depend exclusively on the simulations allows us to not recompute the whole tree at all parameter guesses.

C.3 Unobserved Characteristic

Our goal is to maintain the relatively parsimonious tree representation presented above, while allowing for an unobserved characteristic that may be arbitrarily correlated with the observable ones. Because the covariance between observed characteristics can be estimated in a first stage, it is kept constant through the estimation of the structural parameters. In contrast, the covariance parameters relating the unobserved characteristic with the observed ones are unknown and need to be estimated simultaneously with the preference parameters. As the guesses for the covariance parameters change, so will the simulations of the unobserved characteristic, as in usual simulation-based estimation models.

One complication that arises when incorporating an unobserved characteristic in the context of state-partitioning trees is that the probability of two draws of a continuous variable
coinciding is equal to zero. This uniqueness means that searching the unobserved characteristic not only informs the DM about it, but it also informs the DM of all of the other search characteristics perfectly. Consider the example in Table 19. Unless the cost of evaluating characteristic \( \nu \) is prohibitively high, the DM should opt to immediately learn \( \nu \), because that decision will also provide full information about the remaining characteristics (e.g., learning that \( \nu = 0.3 \) informs the DM that she is in case 1) with certainty.

### Table 19: Example of Simulations with Continuous Unknown Characteristic \( \nu \)

<table>
<thead>
<tr>
<th>Simulation #</th>
<th>( x_1 )</th>
<th>( x_2 )</th>
<th>( \nu )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.3</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>-1.3</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>0.6</td>
</tr>
</tbody>
</table>

An apparent solution would be to discretize the space of \( \nu \), just as one can do with continuous observable characteristics. However, that discretization procedure will fail to update the likelihood for small optimization steps. This is a common issue, usually present in the context of accept/reject estimators (see McFadden (1989)).

We assume that, when consumers take the search action corresponding to the unobserved characteristic, they learn a signal \( \tilde{\nu} = \nu + \eta \), where \( \eta \) is normally distributed \( N \left( 0, \sigma^2_\eta \right) \). For estimation purposes, signal \( \tilde{\nu} \) has a fixed support, based on the sparse grid proposed by Heiss and Wunschel (2008). In this way, the number of support points for \( \nu \) grows with the number of vehicles in the consumer's consideration set. Given the covariance matrix \( \Omega \), which includes the observable characteristics plus the unobserved characteristic \( \nu \), it is possible to write the distribution of \( \tilde{\nu} \) conditional on the observable characteristics. Given this, for a specific parameter guess, we take draws of \( \tilde{\nu} \), conditional on a simulation’s observed characteristics, from the fixed support. The simulation probability of each point \( \tilde{\nu}_i \) is given by \( \tilde{P}_R (\tilde{\nu}_i | x_1, x_2) = \frac{f_{\tilde{\nu}} (\tilde{\nu}_i | x_1, x_2)}{\sum_{j=1}^{NS} f_{\tilde{\nu}} (\tilde{\nu}_j | x_1, x_2)} \) where \( f_{\tilde{\nu}} (\tilde{\nu}_i | x_1, x_2) \) is the density of the signal conditional on the observable characteristics, and \( NS \) is the number of sparse grid nodes used. As \( NS \) grows, our simulation probability \( \tilde{P}_R (\tilde{\nu}_i | x_1, x_2) \) approaches the exact probability \( P_R (\tilde{\nu}_i | x_1, x_2) \) induced by the normal distribution.

Table 20 provides an example. Suppose that only three draws were taken, and the sparse grid of \( \tilde{\nu} \) is \( \{-1.5, 0, 1.5\} \), with simulation probabilities as defined above. According to Table 20, searching the unknown characteristic \( \nu \) partitions the simulation set into subsets \( \{1, 3\} \).
Table 20: Example of Simulations with Continuous Signal $\nu$

<table>
<thead>
<tr>
<th>Simulation #</th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$\nu$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.5</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1.5</td>
</tr>
</tbody>
</table>

and $\{2\}$. The expected value of $\nu$ can also be calculated easily. Before any search activities take place, the DM is in an initial partition $P_1$, and the expected value of $\nu$ is calculated according to

$$E(\nu|P_1) = \frac{1}{3} [E(\nu|x_1 = 1, x_2 = 1, \nu = 1.5) + E(\nu|x_1 = 1, x_2 = 2, \nu = 0) + E(\nu|x_1 = 2, x_2 = 2, \nu = 1.5)]$$

However, if the consumer has searched the unobserved characteristic, and has found that $\nu = 1.5$ such that her partition is $P_2 = \{1, 3\}$, the expected value of the unobserved characteristic equals

$$E(\nu|P_2) = \frac{1}{2} [E(\nu|x_1 = 1, x_2 = 1, \nu = 1.5) + E(\nu|x_1 = 2, x_2 = 2, \nu = 1.5)]$$

The utility from the observed characteristics is calculated similarly.

As before, the simulation draws over $\nu$ may not change across small optimization steps. However, the conditional expectations above do change continuously with the covariance parameters, meaning that this approach does produce a likelihood function that is sensitive to small changes to parameter changes. In our application, we use 30 simulations per consumer. Some characteristics are ex-ante known by consumers, and so, they are always conditioned on, except in counterfactual analyses.

$^{25}$Note also that the loss in efficiency from not updating the simulations of $\nu$ continuously with the covariance parameters decreases as the number of simulations increases.
C.4 Multiple Alternatives

The examples above apply to a single alternative. When the DM has multiple alternatives, it suffices to repeat the same process for each. While this increases the number of columns in Table 20, for example, the number of rows remains at $R$. Clearly, estimator efficiency requires the number of simulations $R$ to be appropriate for the maximum number of alternatives observed in the sample. In our estimation, we cap the sample to consumers who browsed at most 4 vehicles, which comprises 89% of the sample, and we employ 30 simulations per consumer.

Because each row of a simulation table has information about multiple alternatives, it follows that learning a specific vehicle’s characteristic implies partitioning also the simulations for the remaining vehicles, despite the fact that the distribution of characteristics is assumed to be independent across vehicles. Importantly, because the draws are taken independently across vehicles, partitioning the simulation set around a search characteristic of vehicle $i$ has no bearing on the distributional properties of the simulations of the remaining vehicle. Overall, our method ensures that as long as the number of simulations is appropriate, it yields a consistent estimator.

C.5 Calculating the Likelihood of a Search Path

Once the decision-tree is created, according to the steps outlined in Section C.2, it remains to calculate the probability the consumer’s search sequence. Consider a customer who, in the data, searches over a unique alternative with search characteristics $x_1 = 1$ and $x_2 = 2$, and suppose four simulations are drawn, according to Table 21:

<table>
<thead>
<tr>
<th>Simulation #</th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$\tilde{\nu}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.5</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1.5</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>2</td>
<td>1.5</td>
</tr>
</tbody>
</table>

In this example, simulations 1 and 4 are consistent with the vehicle observed in the data. In order to calculate the likelihood of the search path, the algorithm calculates the
probability of each of the customer’s actions, conditional on simulation 1 being true first, and then again conditional on simulation 4 being true. These probabilities are readily available from applying the logit formula to the terminal utilities as well as to the expected utilities of the search actions at every node. Because each simulation was drawn only once, the likelihood of the path can be calculated by a simple average of each of the conditional probabilities. The drawn signals $\tilde{\nu}$ reflect the simulation probabilities, and so no additional adjustments are required.

A potential issue with the current approach is that the drawn simulations may fail to be consistent with the characteristics of the vehicles in the consumer’s consideration set. For example, a vehicle’s characteristics may be $x_1 = 1$ and $x_2 = 2$, while all simulations may predict different levels. This would be an issue, because somewhere along the decision tree, the researchers know that the consumer has learned $x_1 = 1$, which would be incompatible with all of the remaining partitions. In order to solve this problem, we include a single observation in the simulation set that is consistent with the observable characteristics of the vehicles in the consumer’s consideration set. This ensures that there is always a non-empty partition consistent with the consumer’s information set. This step does not affect the consistency of our estimator, because there is no need to increase the number of such simulations beyond one, as $N$ or $R$ increase.
References


64


