Guiding Consumers through Lemons and Peaches: An Analysis of the Effects of Search Design Activities

Pedro M. Gardete* Megan H. Antill†

January 2020

Abstract

The advent of the Internet age has led to questions about the impact of obfuscation and information provision activities by sellers. We estimate the welfare effects of search design activities by an online used car seller, through a dynamic model of search over differentiated offerings. We find that different emphases on product characteristics have a modest impact on consumer welfare (-0.4% to +2.5%) but a relatively high impact on the seller's (-1.9% to +11.3%). Incentives for information provision are found to be largely aligned between supply and demand. Finally, we find no evidence of consumer myopia in learning.

*Nova School of Business and Economics, R. Holanda n.1, 2775-405 Carcavelos, Portugal; Email address: pedro.gardete@novasbe.pt.

†Graduate School of Business, Stanford University, Stanford, California 94305, USA; Email address: mkhunter@stanford.edu.

1 Introduction

Since at least Stigler (1961), economists have considered the effects of search frictions in markets. The onset of the Internet, with its partially-fulfilled promise of complete price transparency, promoted more directed research efforts, including understanding sources of price dispersion, information acquisition strategies by agents, and information obfuscation activities by sellers (Baye, Morgan, and Scholten 2004; Hortaçsu and Syverson 2004; Ellison and Ellison 2009). A fundamental aspect remains understudied. While offering information relevant to consumers, sellers are required to organize how the data are to be supplied. In the the modern Internet era, virtually all sellers are required to engage in so-called 'search design' activities, by which they decide which information to emphasize, and which to relegate.

Sellers of differentiated products face a fundamental tension when engaging in search design. On one hand, emphasizing more informative data is likely to increase consumer welfare. On the other hand, sellers may become better off if they are strategic about how they provide information. In this paper, we empirically investigate to extent to which search design policies affect market outcomes.

Consider the following example, in which a used-car seller offers up 4 vehicles, each with two characteristics:

	Vehicle A	Vehicle B	Vehicle C	Vehicle D
x_1	+20	+20	-20	-20
x_2	-10	+10	-10	+10
$u = x_1 + x_2$	+10	+30	-30	-10

Suppose one of these vehicles is in a given consumer's consideration set, with uniform probability (equivalently, there may exist four consumers, and each searches a different alternative). Under perfect information, consumers derive conversion utility of $u_j = x_1^j + x_2^j$.

The first two rows of the table above denote the vehicles' characteristics, and the last row denotes the consumers' utilities under perfect information. In this example, complete ignorance means that the consumer expects zero utility, because each characteristic is exactly offset by another to provide an ex-ante utility of zero. The consumer can spend a search effort c to learn a characteristic. If the consumer decides to search characteristic 1, she updates her utility to -20 or 20, depending on the vehicle she is inspecting. If instead she searches characteristic 2, her utility is updated to -10 or 10. If the consumer searches both characteristics, then she will find herself in one of four possible distinct cases, each with equal probability. Finally, suppose there exists an outside option - for instance, the current mode of transportation - with known value of 15.

If the firm provides perfect information, then only vehicle B is sold. However, the seller may be able to induce purchases of vehicles A and B by carefully selecting which characteristics to feature. Suppose the seller decides to emphasize characteristic x_1, but not x_2, which now needs to be searched to be learned. In this case, both vehicles A and B (at least momentarily) yield high enough utility to be bought. Knowing x_1, the consumer decides to search x_2 if and only if the search cost is low enough:

$$\begin{aligned} & -c + \underbrace{Pr\left(x_2 = -10 \middle| x_1 = 20\right)}_{Vehicle\,A} \underbrace{15}_{u_0} + \underbrace{Pr\left(x_2 = 10 \middle| x_1 = 20\right)}_{Vehicle\,B} \underbrace{30}_{u_B} \geq \underbrace{20}_{E(u|x_1)} \\ \Leftrightarrow & c \leq 2.5 \end{aligned}$$

When $c > 2.5$, the consumer prefers not to learn x_2, but is willing to buy if she observes $x_1 = 20$ (i.e., vehicles A and B), which occurs with 50% probability: By emphasizing x_1, the seller creates a 'demand wedge' that is willing to purchase a vehicle before acquiring full information.

The example above illustrates that the seller's search design policy may be at odds with the consumers' interests. By disclosing some information the seller generates interest - the market fails with no information - but by shading some other - characteristic x_2 in the example above - the seller uses search costs to stop at least some consumers at the information set with the highest demand level. The example also illustrates that optimal information design is challenging to characterize based on simple heuristics or observations, especially once one considers additional aspects such as correlated product characteristics and heterogeneous consumers.

The goal of this paper is to assess the effects of search design policies. We take advantage of a unique dataset in which we observe consumers' search actions for specific characteristics of vehicles of an online used car seller. The variation from our case study enables us to recover the fundamentals driving consumer learning processes, and conduct counterfactual scenarios of different information disclosure policies by the seller.

We estimate a model of search in which consumers engage in strategic learning. While doing this, they understand that the product characteristics they seek to learn may be arbitrarily correlated. The model introduces some novel features. First, it allows for piecemeal search, which allows us to rationalize our data and conduct counterfactual analysis with different information scenarios. Second, it recovers consumer beliefs flexibly, allowing us to compute consumer expectations at arbitrary information sets. Third, it accounts for an unobservable utility component that is allowed to covary with the observed components, and whose distribution can be identified from search data.

Modeling search is often associated with large state spaces, because researchers are required to allow a large number of outcomes from learning. We develop a simulation-based estimator that takes draws from the distribution of characteristics in order to inform consumer beliefs. As the number of simulations used increases, the simulated actions approximate those of an agent with exact beliefs over the distribution of characteristics.

Our estimation results document preference and search cost parameter estimates, as well as a covariance matrix that summarizes a potentially complex relationship between product characteristics. We document the relative variance in utility explained by each featured characteristic. Our counterfactual analyses reveal that the welfare directions of search design policies tend to be aligned between seller and consumers. Different information design policies are found to affect consumers only moderately, but can have a sizable impact on seller profits: They are found to vary consumer welfare from -0.4% to 2.5%, whereas sales vary between -1.9% and 11.3%.

Past work has analyzed the role of consumer myopia on learning and information provision policies (e.g., Gabaix and Laibson, 2006). Informed by it, we re-estimate the model under the assumption of consumer myopia, and find no evidence of this kind of behavior: All measures of fit worsen - despite the number of parameters being held constant across models - and the myopic model is rejected as an approximation of forward-looking behavior (see Frazier, Powell, and Dayanik (2009), Powell (2010), and Liang, Mu, and Syrgkanis (2017)).

Our paper is related to three important literature streams on search design, search frictions, and myopic consumer learning. The first stream on search design has considered the focal role of price obfuscation (Ellison and Ellison (2009), Einav, Knoepfle, Levin, and Sundaresan (2014)) as well as the role of information in rich two-sided market interactions (Horton (2014), Dinerstein, Einav, Levin, and Sundaresan (2018), Fradkin (2019)). Empirical contributions to the second stream include Hong and Shum (2006), De los Santos, Hortaçsu, and Wildenbeest (2012), and Koulayev (2014). Finally, the third stream (Gabaix and Laibson, 2006, Frazier, Powell, and Dayanik (2009), Powell (2010), and Liang, Mu, and Syrgkanis (2017)) studies the conditions under which myopic policies affect information provision and approximate forward-looking learning behaviors.

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.com, between February and September 2016. 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). Regarding each vehicle, users can 1) browse through the photo gallery, 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.[1] 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.[2] Given the broad product line offered by the firm, we focus on the search data related to browsing activity of sedan vehicles.[3]

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, 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 significantly, 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.

[1]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 to be captured inconsistently in the dataset. We also eliminated consumers whose modal device was a phone or a tablet. We take additional data cleaning steps in order to eliminate irregular browsing activity, as we describe in Appendix A.

[2]The firm provides information on the vehicle prices on its website, which cannot be bargained over. The underlying assumption in using test drives as the conversion variable is that we assume the post test drive behavior remains constant in counterfactual scenarios. Given the high conversion rates, we only model consumer behavior up to the first conversion.

[3]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.

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 website offers a number of search filters that allow consumers to easily identify the vehicles they are willing to consider. Given the low search costs involved at that point, we inform consumers' consideration sets by the vehicles they interact with in the data. Hence, the variation in our data identifies the extent to which consumers search characteristics of alternatives in their consideration sets, rather than the formation of the consideration sets themselves. We believe this is not a primary concern, since the search costs involved in vehicle discovery are low, due to the number of available filters and search functionalities. Despite this, we maintain the prominence of vehicle make-model and color information across counterfactual scenarios which, given their popularity during vehicle discovery activities, reduce concerns related to varying consideration sets further. We focus on the 89% of consumers who evaluated up to four sedan vehicles in the data, yielding a working sample with 12,887 consumers (91% browse up to five vehicles).

We model heterogeneity by incorporating a latent class distribution over search cost levels, such that some consumers may be willing to search certain aspects more than others.[4] Moreover, each segment is allowed to have different search costs related to each information-gathering activity.

2.2 Vehicle Characteristics

We 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.[5]

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

[4]The AIC and BIC criteria favor a two-segment model, which is our model of choice throughout the paper. Details on model comparisons are available from the authors. While extending heterogeneity to preference parameters is trivial, we refrain from doing so due to computational costs, and the fact that none of our counterfactual analyses take advantage of heterogeneity through targeting activities. We detail the computational details of the estimation in Appendix D.

[5]Inspection reports feature one inspection note for each issue found by the dealer.

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 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 their 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.

3 Model

3.1 Utility under Perfect Information

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

$$v_{ijt} = \sum_{k=1}^{K} \beta_k x_{jk} + \nu_{ij} + \epsilon_{ijt} \tag{1}$$

and is equal to

$$v_{i0t} = \epsilon_{i0t} \tag{2}$$

if she decides for the outside option. Above, j indexes vehicles and k indexes observable (by the researchers) vehicle characteristics. Characteristics $\{x_{jk}\}$ are constant across consumers for the same vehicle (e.g., price, mileage, etc), and may be correlated. For example, as discussed in the descriptive analysis, it is reasonable to expect an older vehicle will exhibit higher mileage. Component ν_{ij} is a preference for some vehicle j, which may be correlated with the observed characteristics $\{x_{jk}\}$. The introduction of the unobserved term ν_{ij} captures the fact that, during search, consumers may learn more than the objective characteristics that the researchers are able to observe directly in the dataset. Including term ν_{ij} takes into account that a search action - in our application, photo browsing - may reveal subjective information that is not only challenging to quantify, but also whose judgment may vary across users.[6] Shock ϵ_{ijt} is uncorrelated across consumers, vehicles, and time, and captures

[6]One can also consider decomposing ν_{ij} as $\beta_i.x_j^{'}$, where $x_j^{'}$ is an unobserved product characteristic. We keep notation ν_{ij}, given the impossibility of estimating the distribution of preferences separately from the level of the unknown product characteristic $x_j^{'}$.

idiosyncratic factors independent of the vehicles' characteristics.

The seller's 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 3 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 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 a vehicle's photo gallery provides information about component ν_{ij}.

3.2 Search Dynamics

During their search, 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. In order to take this into account, we assume that agents condition unknown features 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 $1..J_i$, 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. Moving towards a Bellman equation framework, we omit time subscripts, and assume the information-acquisition actions are perturbed by shocks denoted as ϵ_{ij}^{s}, where $s \in S$ denotes a search action. These shocks are learned contemporaneously, and represent unobservable influences that may randomly affect consumers' actions. We denote the collection of user i's shocks as ϵ_i.

We summarize the characteristics known by consumer i about vehicle j by set Ω_{ij}, and denote the collection of all information known about all vehicles in consumer i's consideration set by Ω_i. For example, before initiating search for specific vehicles, user i is endowed with information $\Omega_{ij} = \{make\_model_j, price_j, mileage_j, age_j, color_j\}$ about vehicle j. The information set is then expanded according to the consumer's search decisions.

Decision-Making. At each decision point, consumer i solves the problem:

$$V^i\left(\Omega_i, \epsilon_i\right) = \max \left\{ \underbrace{V_0^i\left(\Omega_i, \epsilon_{i,1}\right), ..., V_{J_i}^i\left(\Omega_i, \epsilon_{i,J_i}\right)}_{Terminal\ Decisions}, \underbrace{V_{1,s}^i\left(\Omega_i, \epsilon_{i,1,1}\right), ..., V_{J_i,s}^i\left(\Omega_i, \epsilon_{i,J_i,s}\right)}_{Search\ Decisions} \Bigg|_{s=1..S} \right\} \tag{3}$$

where $V^i(\cdot)$ is the value function for consumer i.[7] 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. $V_j^i(\cdot)$ denotes the expected value of stopping search and selecting one of the $J_i + 1$ alternatives, and $V_{j,s}^i(\cdot)$ 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

$$V_j^i\left(\Omega_i, \epsilon_i\right) = E\left(\upsilon_{ij} \middle| \Omega_i, \epsilon_i\right) \tag{4}$$

$$= E\left(\sum_{k=1}^{K} \beta_k x_{jk} + \nu_{ij} \middle| \Omega_{ij}\right) + \epsilon_{ij} \tag{5}$$

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

$$V_{j,s}^i\left(\Omega_i, \epsilon_i\right) = -c_s + E_{\epsilon_i', \omega_{js}}\left(V\left(\Omega_i \cup \omega_{js}, \epsilon_i'\right) \middle| \Omega_i\right) + \epsilon_{ijs} \tag{6}$$

where j and s denote the vehicle and search actions to be maximized over. Vector ω_{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 3 and 4.

We allow consumers to incur different search costs depending on their actions, namely at rates c_s, $s \in \{1, 2, 3\}$. We also introduce heterogeneity in search costs by use of two latent

[7] We use notation $V^i(\cdot)$ to stress that consumers may have different sets of vehicles in their consideration sets, and so will face unique value functions. We eliminate repeated search decisions. Although such actions could be incorporated in the model, they are not the main focus of the analysis. We believe these are likely to be related to consumer memory. In this case, the working assumption for the counterfactual analysis is that consumers will conduct the necessary searches to refresh their memories, if needed. For simplification purposes, our notation omits the fact that consumers do not search the same component of the same vehicle multiple times, such that their action space depends on the state. Future work may inform memory-related parameters by extending our model to allow repeated search action.

segments, which vary in size and in the values of c_s.

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

$$\Omega_{ij} = \{make\_model_j, price_j, mileage_j, age_j, color_j\} \tag{7}$$

to

$$\Omega_{ij}^{'} = \{make\_model_j, price_j, mileage_j, age_j, color_j, \nu_{ij}\} \tag{8}$$

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 $\epsilon_i^{\prime}$ as well as to the information ω_{js} that the consumer will obtain through search, and which is affected by the information in Ω_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 ν_{ij} and ϵ_{ijt}). We denote the multivariate cumulative distribution (c.d.f.) of these characteristics by

$$\mathcal{G}_{make\_model,price,mileage,age,color,owners,accidents,notes} \tag{9}$$

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 $\mathcal{F}_{X_1..X_K}$ admit copula representations

$$C\left(F_{X_1}(x_1), ..., F_{X_K}(x_K)\right) \tag{10}$$

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's color 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 ν_{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 were found in the dataset). We estimate an ordered probit regression for inspection notes according to the model

$$notes_j = \begin{cases} 0, & y_j^* \leq \mu_0 \\ 1, & \mu_0 < y_j^* \leq \mu_1 \\ \vdots & \vdots \\ 6, & y_j^* > \mu_6 \end{cases} \tag{11}$$

where

$$y_j^* = \beta_{mm}^{notes} make\_model_j + \beta_{color}^{notes} color_j + \epsilon_j^{notes}, \tag{12}$$

$make\_model$ and $color$ are indicator variables, and $\epsilon_j^{notes} \sim N(0, 1)$. We employ the same approach to the remaining variables, including price and mileage, which we discretize into five levels each.[8] 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 and color fixed effects. Note that the estimated residuals $\widehat{\epsilon}_j$ are independent of the non-ordinal variables. Because each discrete value of the dependent variables can be generated by several values of ϵ_j, we need to define a specification for estimator $\widehat{\epsilon}_j$. We use the maximum likelihood estimator of $\widehat{\epsilon}_j$, so that each level of each characteristic is associated with a unique level of the estimated residual $\widehat{\epsilon}_j$. By construction, these residuals are independent of the categorical variables $make\_model$ and $color$.[9]

[8] Discretizing these variables into additional levels is straightforward, with a relatively low penalty for our estimator.

[9] An alternative (and more demanding approach for our estimation strategy) would be to associate multiple levels of residuals to each characteristic level, by simulating ϵ_j's from intervals of the ordinal probit regressions.

We denote the joint distribution of the estimated residuals as

$$\mathcal{F}_{price,mileage,age,owners,accidents,notes \perp make\_model,color} \tag{13}$$

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. These variables are introduced into a Gaussian copula:

$$c_{\Sigma}(u) = \frac{1}{\sqrt{\det \Sigma}} \exp \left\{ -\frac{1}{2} \begin{pmatrix} \Phi^{-1}(u_1) \\ \vdots \\ \Phi^{-1}(u_K) \end{pmatrix}^{'} .\Sigma^{-1}. \begin{pmatrix} \Phi^{-1}(u_1) \\ \vdots \\ \Phi^{-1}(u_K) \end{pmatrix} \right\} \tag{14}$$

where

$$u = F(\varepsilon) = \begin{pmatrix} F_1(\varepsilon^1) \\ \vdots \\ F_K\left(\varepsilon^K\right) \end{pmatrix}$$

and where matrix Σ is a correlation matrix (its main diagonal equals one, by definition), 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 $\left(\Phi^{-1}(u_1) \ldots \Phi^{-1}(u_K)\right)^{'}$. 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 random vector $Z = \{z_1 .. z_K\}$:

$$f_{Z_1..Z_K}(z) = \frac{1}{\sqrt{\det \Sigma}} \exp \left\{ -\frac{1}{2} \begin{pmatrix} z_1 \\ \vdots \\ z_K \end{pmatrix}^{'} .\Sigma^{-1}. \begin{pmatrix} z_1 \\ \vdots \\ z_K \end{pmatrix} \right\} \tag{15}$$

where $z_k \sim N(0,1)$. Above, the elements of random vector Z follow non-independent standard normal distributions, with covariance matrix equal to Σ, which together with the marginal distributions characterizes the joint distribution of the ordinal characteristics.

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

First, the residuals of each characteristic $\widehat{\varepsilon}^k$, orthogonal to make-model and color information, are estimated through the transformation

$$\widehat{\varepsilon}_k = G_k^{-1}(y_k, X) \tag{16}$$

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

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

$$
\begin{tikzcd}
x_j \arrow[r, bend left, "G_j^{-1}"] & \varepsilon_j \arrow[r, bend left, "F_j"] & u_j \arrow[r, bend left, "\Phi^{-1}"] & z_j \arrow[r, phantom, "\backsim"] \arrow[r, bend right, "\Sigma"'] & f_z
\end{tikzcd}
$$

indicator variables with make-model and color information for each vehicle. Calculating the empirical percentiles of ε_k for each k (or in different words, evaluating the empirical distributions F_k at values ε_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, Σ 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, for simulation during estimation.

It remains to solve the second estimation challenge, which is to include the unobserved utility component ν_{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:

$$
f_{Z_1..Z_K,\nu}(z) = \frac{1}{\sqrt{\det \Sigma_0}} \exp \left\{ -\frac{1}{2} \begin{pmatrix} z_1 \\ \vdots \\ z_K \\ \nu \end{pmatrix}' \cdot \Bigg(\underbrace{\begin{matrix} \begin{bmatrix} & & \\ & \Sigma & \\ & & \end{bmatrix} & \sigma_{z\nu} \\ \sigma'_{z\nu} & \sigma^2_\nu \end{matrix}}_{\Sigma_0} \Bigg)^{-1} \cdot \begin{pmatrix} z_1 \\ \vdots \\ z_K \\ \nu \end{pmatrix} \right\} \tag{17}
$$

The covariance matrix Σ is now bordered by the relationships between the unobserved component ν and each of the observable attributes, making up matrix Σ_0. Vector $\sigma_{z\nu}$ contains the covariances between ν and each 'normalized characteristic' z_k, and σ^2_ν is a scalar corresponding to the variance of ν. This specification allows us to identify correlations between observed characteristics and the unobserved utility component.

3.4 Estimation

Estimation is complicated by the size of the state space, part of which is unobservable. In addition to their 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 ω 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 potential elements:

$$|\Omega_{ij}| = \underbrace{(1+7)}_{insp.\,notes} \times \underbrace{(1+4\times 6)}_{vehicle\,hist.} \times \underbrace{(1+\omega)}_{photos} = 88\,(1+w) \tag{18}$$

where the '$(1+\cdot)$' 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

$$|\Omega_i| = J_i^{88(1+w)} \tag{19}$$

For example, a consumer with four vehicles in her consideration set, and with unobserved characteristics approximated with $w = 5$ points, will face approximately 7.72×10^{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 may not overlap exactly).

In order to estimate our model, we develop an estimator that relies on simulating beliefs. Consumer beliefs are replaced by draws from the distribution of product characteristics. Consumers are then assumed to know the realized draws, but are unaware of which corresponds to the products they are searching. Importantly, as the number of simulations increases, the simulated beliefs converge to their population analogue.

We refer the reader to Appendix C for the precise estimator details. An intuitive explanation for our estimator is that, rather than writing the likelihood of fully rational consumers - who are able to keep an incredible number of possibilities in mind - we instead estimate our model as if the likelihood was informed by the actions of boundedly-rational consumers, whose beliefs are as sophisticated as defined by the number of simulations used during estimation. Importantly, the continuous mapping theorem (Mann and Wald, 1943) implies that our estimates converge in distribution to that of agents with perfect beliefs, as the number of simulations increases.[10]

Our approach has a number of advantages. First, it is scalable: It is possible to represent

[10] In the appendix, we also present a sketch of the pseudocode necessary to represent the corresponding value functions.

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 samples more from higher probability areas more often, such that the approximation proceeds by prioritizing high-probability events. This produces an arguably near-ideal approximation sequence. We now provide an outline of the estimation procedure.

First Stage Estimation Steps

- First, estimate matrix Σ (correlation of observable characteristics), according to the steps laid out in Section 3.3. (Note that the additional border that makes up matrix Σ_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 each consideration set, conditional on the vehicle's ex-ante characteristics. Refer to Figure B below for the steps involved in this simulation, discussed later.
- 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.

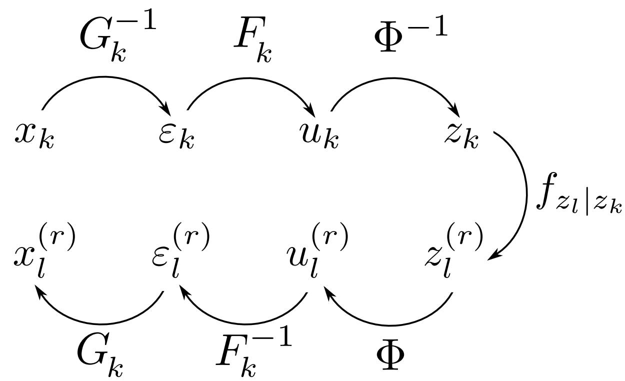
Second Stage Estimation Steps

- Select the initial values for the preference parameters, the search cost parameters, and the covariance parameters $\sigma_{z\nu}$. (We later explain that σ_ν^2 is normalized to one.)
- Draw values of ν conditional on $\sigma_{z\nu}$.
- For each consumer, update the value function by calculating utilities backwards from perfect information.
- Use the decision tree to calculate the joint probability of the consumer's search and purchase decisions observed in the data, for each simulation of the unobserved component.
- Average the probabilities across simulations and sum them across consumers and calculate the log-likelihood for the sample. Decide convergence and/or attempt new parameter estimates.

Figure 6 depicts a decision tree for a consumer searching over a single vehicle. The tree depicts outcomes of 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. Details are provided in Appendix C.

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 ν, 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.

Figure B: Obtaining draws of characteristics x_l, conditional on characteristics x_k



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

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 *age*, conditional on *price*. First, we convert *price* to z_{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 $\left\{z^r_{age}\right\}_{r=1}^{R}$, conditional on z_{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 are then updated for each parameter guess of $\sigma_{z\nu}$.

Variance-Covariance Matrix. Matrix Σ_0 moderates the relations between the vehicles' observed characteristics, and its border includes the relations with the idiosyncratic utility component ν. As we further discuss in the Identification section, we normalize parameter σ_ν^2 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

$$C'.C = \begin{pmatrix} \begin{bmatrix} & & \\ & \Sigma & \\ & & \end{bmatrix} & \sigma_{z\nu} \\ \sigma'_{z\nu} & 1 \end{pmatrix} \tag{20}$$

The process of finding C follows from common Cholesky decomposition. In appendix B, we explain how we parameterize matrix C in order for Σ_0 to be simultaneously consistent with the correlation matrix of the observable shocks Σ, estimated in the first stage, and for Σ_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

$$L\left(\theta | X, A_i\right) = \Pi_{i=1}^{N} Pr\left(A_i | \theta, X\right) \tag{21}$$

where X contains vehicle characteristic data and the correlation matrix Σ, $A_i = \left\{a_i^1, ..., a_i^{T_i}\right\}$ is the sequence of actions of individual i, T_i is the number of actions taken by individual i, and θ is a vector of parameters. Because the shocks ν_{ij} are unknown to the econometricians, we employ a simulated maximum likelihood approach:

$$\check{P}r\left(A_i | \theta\right) = \frac{1}{R}\sum_{r=1}^{R} Pr\left(A_i | \theta, \nu_i^r\right) \tag{22}$$

and the simulated log-likelihood for the sample follows:

$$\begin{aligned} l\left(\theta | X\right) &= \log\left(\Pi_{i=1}^{N} \check{P}r\left(A_i | \theta\right)\right) \\ &= \sum_{i=1}^{N} \log\left(\check{P}r\left(A_i | \theta\right)\right) \end{aligned} \tag{23}$$

3.5 Identification

The identification of the search model is relatively complex, because most 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 β_1 is higher. The reason is that higher values of β_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.[11]

It remains to discuss the identification of the covariance parameters between the observed characteristics and the unobserved utility component ν. 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 ν. The price threshold at which the consumer is indifferent between continuing her search for ν 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 ν, conditional on the known characteristics. The parameters $\sigma_{z\nu}$ of the variance-covariance matrix thus identify the residual variance of component ν, conditional on knowledge of the vehicles' remaining characteristics.

[11]This normalization applies to virtually all models of search for unobserved components.

4 Results

Table 5 presents the estimates of the preference coefficients. Consumers are found to dislike more expensive vehicles, with higher mileage, higher number of inspection issues, and with a higher number of previous owners. We find no statistically significant effect for vehicle age and the number of previous accidents. This may be explained by the remaining regressors - consumers may deem age irrelevant once mileage is controlled for - or by a lack of power (e.g., very few vehicles exhibit more than one past accident). Overall, the statistical significant parameter estimates are in line with common expectation.

Table 6 presents the estimates of the search cost parameters by segment. The size of the first segment is of approximately 67%. These consumers prioritize learning about vehicle histories to browsing photos and inspection reports. Segment two, on the other hand, privileges photos before anything else. This segment is consistent with the presence of a portion of consumers in the dataset who appear to spend most of their time browsing photo galleries of different vehicles. The search costs estimates are interpreted as a function of the normalized variance of the unknown utility component, σ_ν^2. As the researchers set σ_ν^2 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 σ_ν^2. Similarly, when σ_ν^2 is normalized to a low number, search costs estimates decrease. 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 the parameter values should not be immediately interpreted as monetary search costs, even when normalized by the price coefficient.

Table 7 presents the estimates of the variance-covariance matrix of the unobserved term, which are all significant at the 5% level. Table 8 presents the implied variance-covariance matrix of the observable and unobservable characteristics. The non-bold components of the table represent the correlations across observed characteristics, with the influences of make-model and vehicle color parsed out. They are largely consistent with the raw correlations presented in Table 2. As for the unobservable utility component - learned by inspection of vehicle pictures - we find that it follows a familiar pattern: It correlates negatively with price and positively with the remaining characteristics.

Table 9 provides a comparison between the model predictions and some key moments in the data. Overall, the model matches average search intensities reasonably well, especially given that the model's parameters are estimated via likelihood maximization rather than through matching the moments in the data directly. Nonetheless, the model tends to underestimate the maximum of the search intensities observed in the data.

5 Counterfactual Analyses

5.1 Utility Explained by each Vehicle Characteristic

Combining information about preference parameters and the variance-covariance matrix of characteristics, we simulate how much of utility variation is explained by each characteristic. We calculate the familiar statistic

$$\%\,\text{Utility Variance Explained by}\, x_j = \frac{Var\left(E\left(u|\, x_j\right)\right)}{Var\left(u\right)} \tag{24}$$

which captures how much of the utility variance is explained by each characteristic.

Figure 7 presents the percent utility variance explained by each characteristic, by order of importance. Of the four characteristics found to be statistically-significant (cfr. Table 5), three top the list, namely, the number of previous owners, vehicle mileage, and number of inspection notes. Of note is the fact that price is found to explain little variation in utility. This result is unexpected, as one may expect price to be the decisive characteristic in this market. However, it is a consequence of two separate quantities: First, the preference parameters associated with each characteristic, and second, the amount of variation of the characteristics in the data.[12] In a relatively competitive market - such as the case of the online used car market - prices may indeed affect preferences significantly, but competition may also act as a downward force on actual variation. Our results should therefore be interpreted as the combination of the two forces (preferences and characteristic variation), which appear to make price relatively unimportant in explaining total variation in utility.

5.2 Designing Information Provision

We conduct counterfactual analyses across different information structures. Each information structure emphasizes certain characteristics on the main page, and relegates others to the vehicle detail pages. The 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 seller's website liken a spreadsheet, severely affecting user experience.[13] Each counterfactual analysis recomputes consumer search and purchase decisions. Consumers' expectations change, according to the information they have available on the main listing page, as well as with the information they acquire during their search.

We first exchange characteristics between the two levels, one at a time. Table 10 presents

[12] For completeness, note that this statistic also captures a third factor: The cross-correlations in characteristics.

[13] In the same vein, we do not consider the counterfactual scenario of moving a vehicle's photo gallery to the main listing page.

the consumer welfare effects of such policies. We find that emphasizing vehicle histories has statistically significant positive welfare effects on consumers. On the other hand, emphasizing inspection reports may have negative influences. The relative impact is modest, ranging from -0.41% to approximately 2.5%. The fact that highest welfare increase involves price obfuscation (i.e., prices would only be disclosed in detail pages) is highlighted in the discussion below.

Table 11 presents the effects on sales. These effects are found to be higher than those related to consumer welfare, ranging from -1.91% to +11.34%. In fact, search design is found to always have higher relative impacts on profitability than on consumer welfare, as we show later. Inspection of Table 11 reveals that price obfuscation - in our context, the act of relegating price information to vehicle detail pages - can be especially profitable to the seller. The result of a positive impact of this action on consumer and seller welfare requires further consideration. We first summarize the relationship of all of the consumer and seller effects in Figure 8, including the remaining information scenarios, which we detail in Table 12. This analysis reveals that, across search designs, a very strong and approximately linear relation occurs between consumer and seller effects ($R^2 \simeq 97\%$) across the different information provision scenarios. This finding is surprising and at odds with the view that information design is a zero-sum game between the seller and consumers. Rather, our findings point to overwhelmingly consistent incentives. One reason for this may be that our analysis utilizes variation from a seller in a strongly competitive market. As such, our results are likely to reflect the fact that the seller is better off emphasizing the data that is most informative to consumers. In support of this, note that Tables 10 and 12 reveal that the most profitable scenarios involve the seller emphasizing vehicles' histories. In these are included the vehicles' number of previous owners, which is the most important explanatory variable of utility variability (cfr. Figure 7).

Our finding of positive correlation between seller and consumers' utilities across search designs is at odds, for example, with the spirit of the analysis by Ellison and Ellison (2009). We posit that the differences are likely to be explained by different levels of market power: As competitiveness increases, seller incentives become more aligned with consumers', and in the limit of perfect competition, sellers would only survive if they selected the information provision strategy that maximizes consumer surplus. Future work may be able to empirically document a significant correlation between market power and the alignment of information provision incentives formally.

5.3 Consumer Myopia

We adapt our model to consider consumers who are myopic in terms of the benefits produced by learning. Specifically, we consider a model where consumers act as if they only consider the immediate learning benefit of each search action, but ignore the benefits for subsequent searches at each point (see Frazier, Powell, and Dayanik (2009), Powell (2010), and Liang, Mu, and Syrgkanis (2017)). Under this assumption, terminal utilities are calculated as before:

$$V_j\left(\Omega_i, \epsilon_i\right) = E\left(v_{ij} \middle| \Omega_i, \epsilon_i\right) \tag{25}$$

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

$$\begin{aligned} V_{j,s}\left(\Omega_i, \epsilon_i\right) = &-c_s + \\ &E_{\epsilon_i^{\prime}, \omega_{js}}\left[\max\left\{v_{i0}, V_1\left(\Omega_i, \epsilon_i^{\prime}\right), .., V_j\left(\Omega_i \cup \omega_{js}, \epsilon_i^{\prime}\right), .., V_J\left(\Omega_i, \epsilon_i^{\prime}\right)\right\} \middle| \Omega_i\right] + \epsilon_{ij}^s \end{aligned} \tag{26}$$

The formula above reveals that myopic consumers behave as if each search was their last one.

We adapt our value functions by recalculating search values according to (26), and re-estimate the model. Notice that while utility calculations differ between models, the number of estimated parameters remains constant across models.

We first compare the predicted moments of the myopic model with the ones in the data in Table 13. The fit measures can be compared with those of the full forward-looking case, presented in Table 9. Across the board, the myopic model fits the sample moments worse than the forward-looking model.

Second, 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. This test rejects that the models are at the same distance from the true data generation process with a p-value less than 1% (see Table 14).

Overall, these results suggest that, albeit computationally convenient, the myopic model provides a poor approximation to the forward-looking problem. These findings are in contrast, for example, with the conditions isolated by the work mentioned above, which states that myopic behavior may be at times optimal. The results also question the degree to which consumers are myopic about the payoffs of engaging in search activities. Indeed, in our case, failing to incorporate strategic behavior results in worse fit of the moments in the data and in a statistically-significant different likelihood than of the full dynamic problem.

6 Concluding Remarks

Our analysis has concerned itself with estimating the effects of search design policies on consumer and seller welfares. In their theoretical paper, Ellison and Wolitzky (2012) stated that "retailers may (...) harness the power of the Internet to make information processing problems more formidable [to consumers]." Our results are contrasting with their work, in that we mostly find a coincidence of interests between seller and consumers. We expect that additional research, relating sellers' incentives to obfuscate information with market power, may reconcile the findings.

There is also space to consider the extent to which different disclosure policies affect consumers' beliefs about the distribution of characteristics they are likely to face. This task is challenging because of the complications introduced by the introduction of the Bayesian equilibrium concept, and its constraints on the beliefs over the distribution of characteristics. The approach would nonetheless represent a significant step forward for the empirical literature on search design.

The effects of different information scenarios can also be estimated through the use of experiments. However, each experimental condition requires specific investments, which present real design and roll-out costs, and may not be scalable - we have reported the findings of eleven scenarios in this paper. In addition, sellers are often conservative about providing different users with different experiences. In contrast with experiments, our methodology can be viewed as a way to recover the effects of search design policies from non-experimental revealed-preference data.

7 Figures

Figure 1: Screenshot of Main Listing Page, shift.com, October 2016

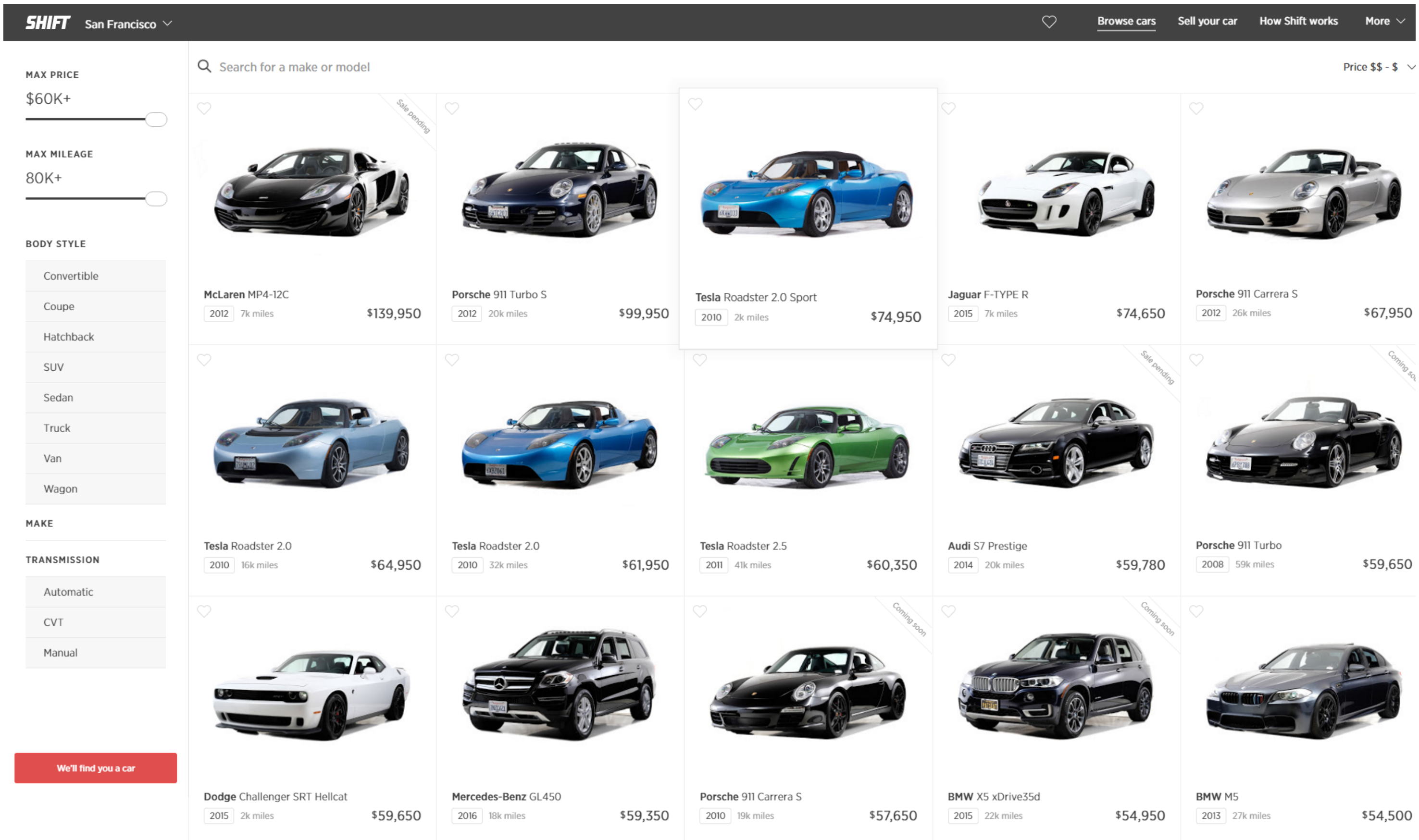


Figure 2: Annotated Screenshot of Vehicle Detail Page, shift.com, October 2016

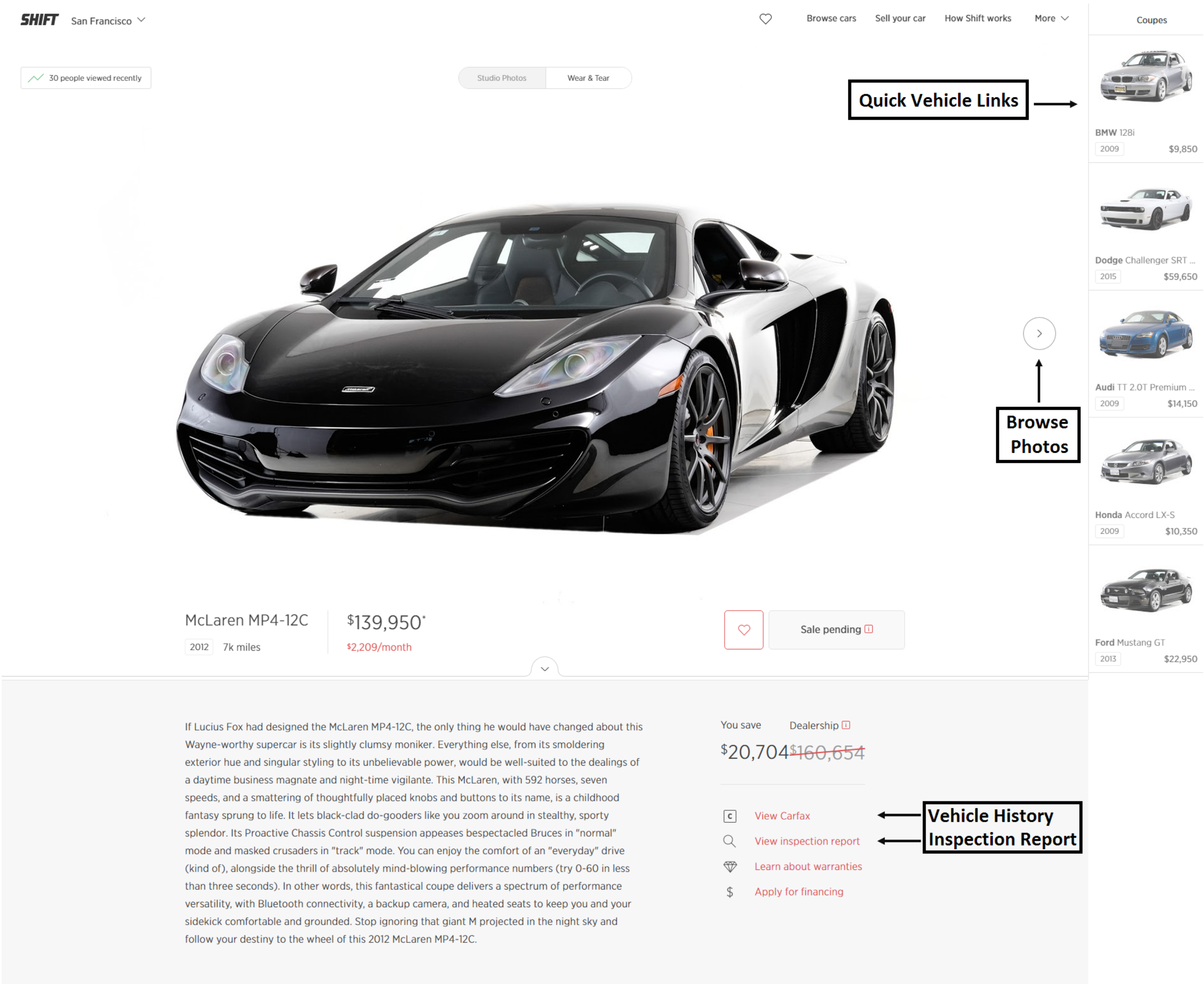
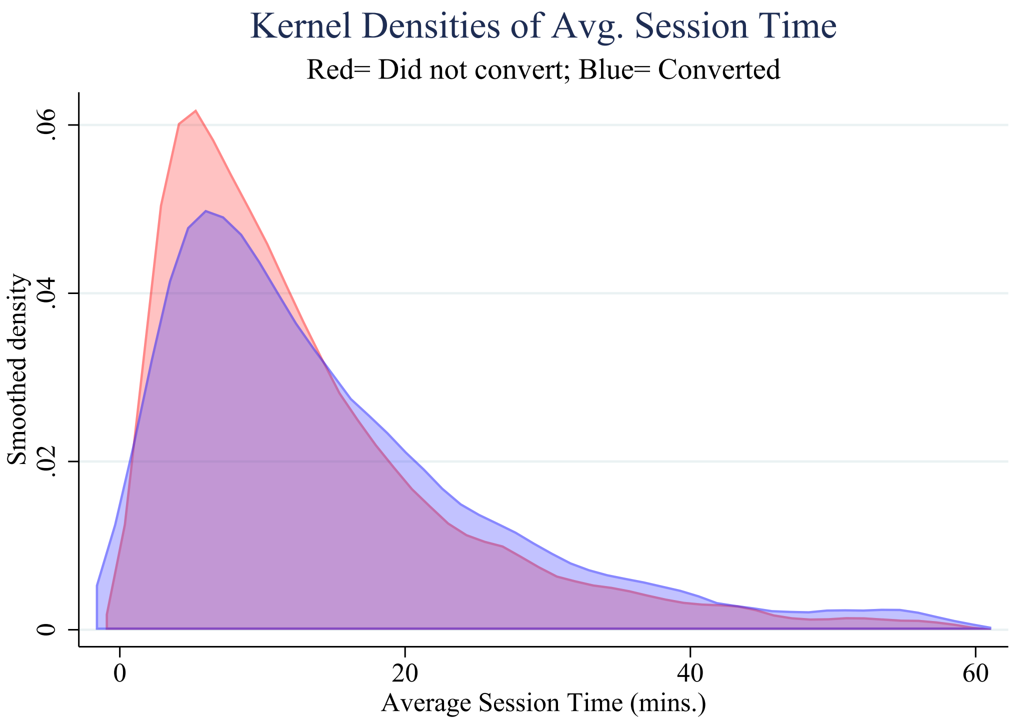
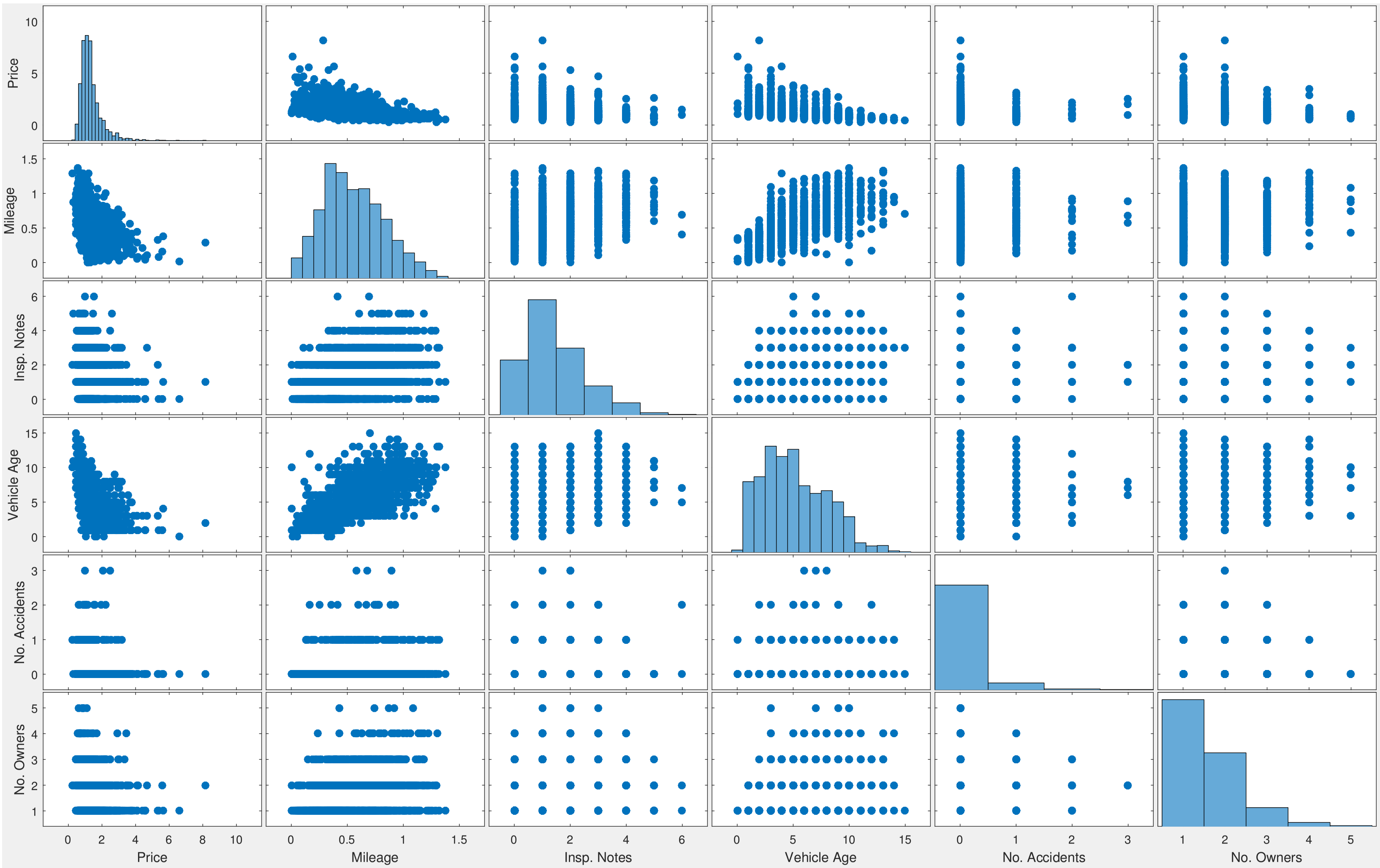


Figure 3: Densities of Avg. Session Times Across Users



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

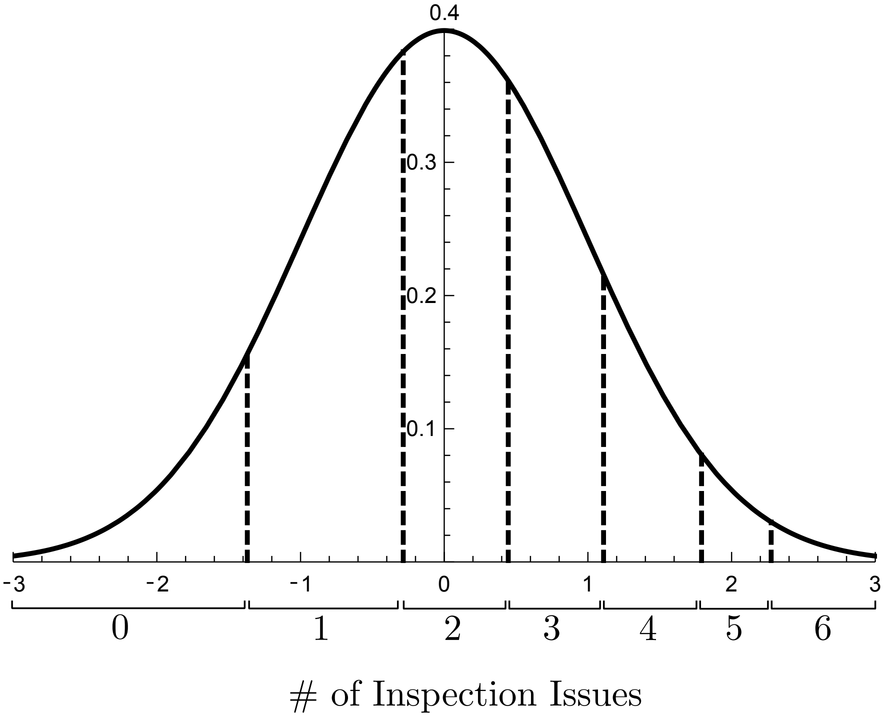
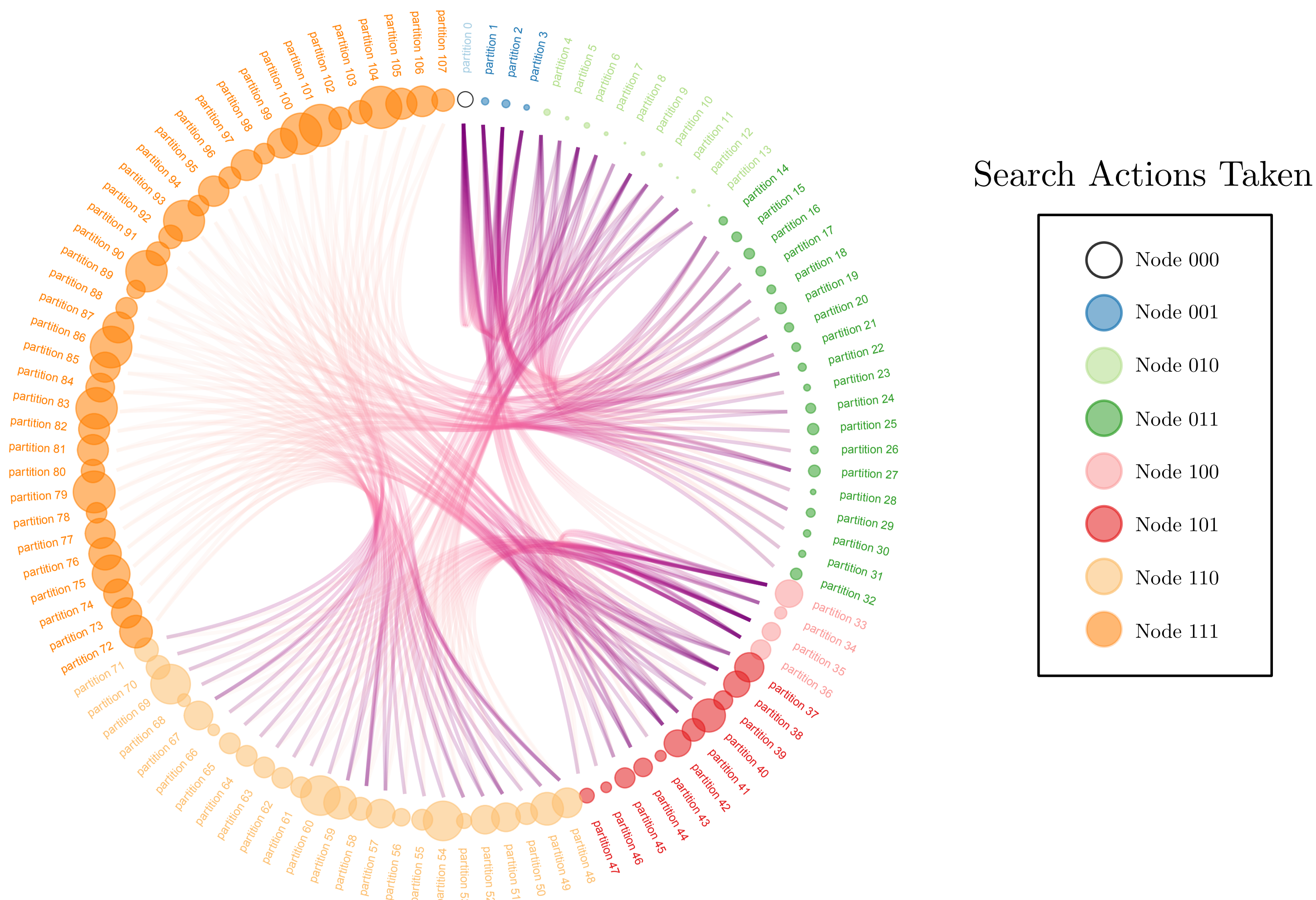
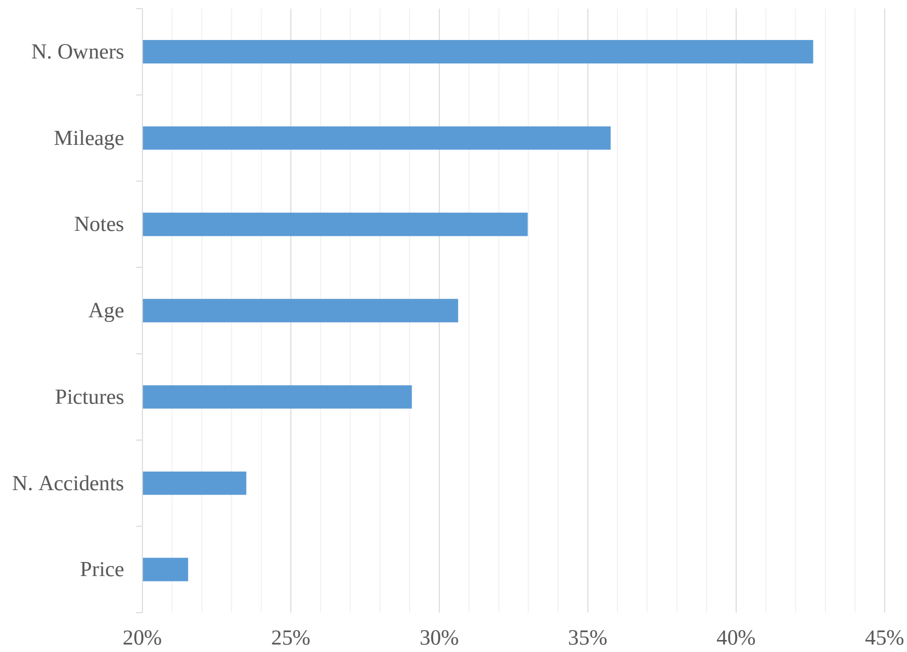


Figure 6: Example of Decision Tree



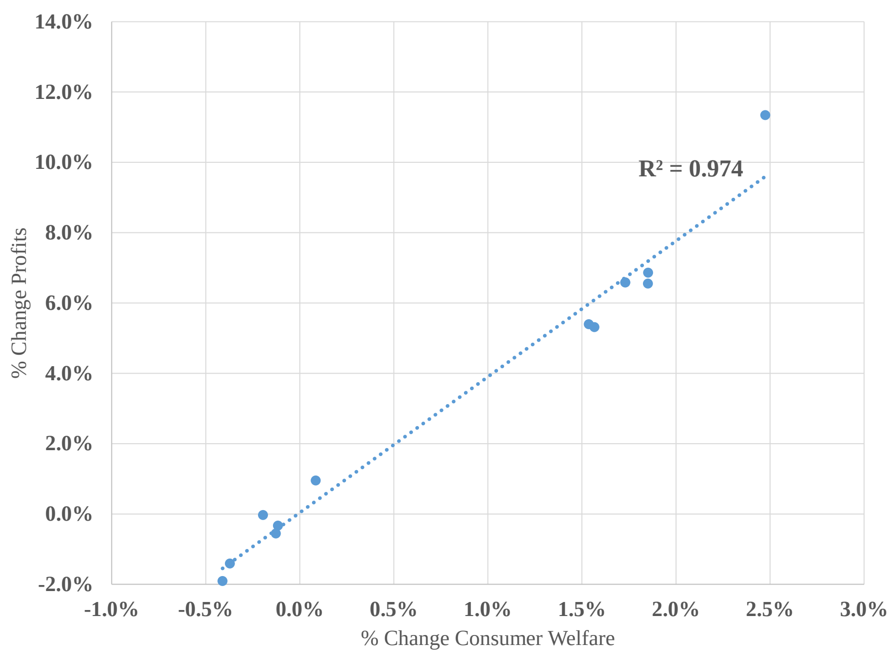
Note: Decision tree, generated for consumer searching for one vehicle, and estimation with 50 simulations.

Figure 7: % Utility Variance Explained by each Characteristic



Note: Above, values obtained by simulating $Var_{x_j}\left[E\left(u|\, x_j\right)\right] \div Var\left(u\right)$.

Figure 8: Information Design Effects: Consumer Welfare and Profits



Tables

Table 1: Descriptive Statistics

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

	Price	Mileage	Insp. notes	Vehicle age	No. accidents	No. owners
Price	1					
Mileage	-0.45	1				
Insp. notes	-0.16	0.39	1			
Vehicle age	-0.40	0.71	0.44	1		
No. accidents	-0.07	0.10	0.09	0.12	1	
No. owners	-0.08	0.21	0.09	0.17	0.09	1

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

Table 3: Vehicle Information and State Variables

Knowledge Level	Characteristic	Utility Component
	Ex-ante Characteristics	
Observable ex-ante	Make/Model	$make\_model_j$
>>	Price	$price_j$
>>	Mileage	$mileage_j$
>>	Age	age_j
>>	Color	$color_j$
	Perceived Quality/Fit	
Learned after browsing one photo set	Vehicle Photos	ν_{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 4: Correspondence of Search Actions and Learned Characteristics

Search Action (s)	Meaning	Learned Characteristics
1	Vehicle Photos	ν_{ij}
2	Vehicle History	$owners_j, acc_j$
3	Inspection Report	$notes_j$

Table 5: Model Estimates: Preference Parameters

	Parameter Estimates
Vehicle Characteristics	
price	-0.372**
	(0.000)
mileage	-0.19**
	(0.000)
notes	-0.36**
	(0.000)
age	-0.01
	(0.348)
accidents	0.032
	(0.434)
owners	-1.611**
	(0.000)
Log-Likelihood: -38,937.35	
Make-Model and Color Dummies ✓	
N= 12,887	

Note: Standard errors in parentheses. Significance levels: † $p \leq 0.10$, * $p \leq 0.05$, ** $p \leq 0.01$.

Table 6: Model Estimates: Search Cost Parameters

	Parameter Estimates
Segment 1	
c_{photos}	2.306**
	(0.031)
$c_{vehicle\,history}$	0.248**
	(0.048)
$c_{insp.report}$	5.255**
	(0.094)
Segment 2	
c_{photos}	0.034*
	(0.02)
$c_{vehicle\,history}$	4.382**
	(0.05)
$c_{insp.report}$	3.309**
	(0.022)
Other Parameters	
α	-0.732**
	(0.01)
Make-Model and Color Dummies	✓
N= 12,887	

Note: Search costs estimated through $\exp(\cdot)$ transformation. Standard errors in parentheses obtained via the delta method. Significance levels: † p≤0.10, * p≤0.05, ** p≤0.01. Size of segment 1, implied by $\widehat{\alpha}$, is 67.5%.

Table 7: Model Estimates: Variance-Covariance Parameters

Variance-Covariance Parameters		Parameter Estimates
s_1		-0.445**
		(0.045)
s_2		0.76**
		(0.000)
s_3		0.674**
		(0.000)
s_4		0.609**
		(0.044)
s_5		0.1*
		(0.045)
s_6		0.551**
		(0.045)
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 8: Estimated Variance-Covariance Matrix

$$\widehat{\Sigma_0} = \left(\begin{array}{ccccccc} 1 & -0.66 & -0.31 & -0.73 & -0.07 & -0.21 & \mathbf{-0.44} \\ -0.66 & 1 & 0.37 & 0.71 & 0.08 & 0.27 & \mathbf{0.76} \\ -0.31 & 0.37 & 1 & 0.41 & 0.07 & 0.14 & \mathbf{0.67} \\ -0.73 & 0.71 & 0.41 & 1 & 0.09 & 0.24 & \mathbf{0.61} \\ -0.07 & 0.08 & 0.07 & 0.09 & 1 & 0.021 & \mathbf{0.1} \\ -0.21 & 0.27 & 0.14 & 0.24 & 0.021 & 1 & \mathbf{0.55} \\ \mathbf{-0.44} & \mathbf{0.76} & \mathbf{0.67} & \mathbf{0.61} & \mathbf{0.1} & \mathbf{0.55} & 1 \end{array}\right) \Bigg| \left(\begin{array}{c} Price \\ Mileage \\ Insp.notes \\ Vehicle\,age \\ No.accidents \\ No.owners \\ \nu \end{array}\right)$$

Note: In bold, cross-correlation elements induced by estimated parameters $s_1..s_6$. Rightmost vector shows the corresponding characteristics. Non-bold figures are all significant at the 1% level. Bold figures are significant at the 5% level (see Table 7).

Table 9: Search Moments from Dataset and Model Predictions

	Data				Model Prediction			
	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
Conversion Rate	0.047	0.213	0	1	0.055	0.228	0	1
N. Searches	1.639	3.22	1.082	9	1.433	1.153	0.5	6.054
Vehicle Histories	0.441	0.743	0	4	0.397	0.316	0.108	1.844
Inspection Reports	0.175	0.496	0	4	0.161	0.184	0	1.157
Photo Sets	1.023	0.873	0	4	0.875	0.664	0.268	3.376
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 10: Consumer Welfare Effects of Exchanging Attributes Between Front and Detail Pages

	Vehicle Detail Page Attribute	
Front Page Attribute	Vehicle History	Inspection Report
Mileage	+1.65% (+1.73%)**	-0.39% (-0.41%)**
Age	+1.76% (+1.85%)**	-0.11% (-0.12%)
Price	+2.35% (+2.48%)**	+0.02% (+0.01%)*
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. Relative changes in parentheses. Significance levels: † p≤0.10, * p≤0.05, ** p≤0.01.

Table 11: Conversion Effects of Exchanging Attributes Between Front and Detail Pages

	Vehicle Detail Page Attribute	
Front Page Attribute	Vehicle History	Inspection Report
Mileage	+0.36% (+6.58%)**	-0.11% (-1.91%)
Age	+0.36% (+6.55%)**	-0.02% (-0.33%)
Price	+0.63% (+11.34%)**	-0.00% (-0.03%)
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. Relative changes in parentheses. Significance levels: † p≤0.10, * p≤0.05, ** p≤0.01.

Table 12: Conversion Effects of Different Information Configurations

	Information Configuration		Consumer Welfare	Conversion Effects
Original Configuration:	Main Listing Page	Vehicle Detail Page	–	–
	$\begin{bmatrix} Mileage \\ Age \end{bmatrix}$	$\begin{bmatrix} V.Hist. \\ Insp.Rep. \end{bmatrix}$		
Scenario a)	$\begin{bmatrix} V.Hist. \\ Insp.Rep. \end{bmatrix}$	$\begin{bmatrix} Mileage \\ Age \end{bmatrix}$	+1.49% (+1.57%)**	+0.29% (+5.31%)**
Scenario b)	$\begin{bmatrix} V.Hist. \\ Insp.Rep. \end{bmatrix}$	$\begin{bmatrix} Age \\ Mileage \end{bmatrix}$	+1.46% (+1.54%)**	+0.30% (+5.39%)**
Scenario c)	$\begin{bmatrix} Mileage \\ Age \end{bmatrix}$	$\begin{bmatrix} Insp.Rep. \\ V.Hist \end{bmatrix}$	+0.08% (+0.09%)	+0.05% (+0.95%)
Scenario d)	$\begin{bmatrix} Insp.Rep. \\ Age \end{bmatrix}$	$\begin{bmatrix} Mileage \\ V.Hist \end{bmatrix}$	-0.35% (-0.37%)**	-0.08% (-1.41%)
Scenario e)	$\begin{bmatrix} Mileage \\ Insp.Rep. \end{bmatrix}$	$\begin{bmatrix} Age \\ V.Hist \end{bmatrix}$	-0.12% (-0.13%)†	-0.03% (-0.56%)
Scenario f)	$\begin{bmatrix} V.Hist \\ Age \end{bmatrix}$	$\begin{bmatrix} Insp.Rep. \\ Mileage \end{bmatrix}$	+1.65% (+1.73%)**	+0.36% (+6.58%)**
Scenario g)	$\begin{bmatrix} Mileage \\ V.Hist \end{bmatrix}$	$\begin{bmatrix} Insp.Rep. \\ Age \end{bmatrix}$	+1.76% (+1.85%)**	+0.38% (+6.86%)**
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. Relative changes in parentheses. Significance levels: † p≤0.10, * p≤0.05, ** p≤0.01.

Table 13: Consumer Behavior Statistics Compared with Myopic Model

	Data				Knowledge Gradient			
	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
Conversion Rate	0.047	0.213	0	1	0.057	0.233	0	1
N. Searches	1.639	3.22	1.082	9	1.16	0.65	0.46	4.319
Vehicle Histories	0.441	0.743	0	4	0.279	0.163	0.057	1.143
Inspection Reports	0.175	0.496	0	4	0.119	0.084	0	0.694
Photo Sets	1.023	0.873	0	4	0.762	0.428	0.257	2.699
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 14: Vuong's Non-Nested Test

Log-likelihood		Vuong's Test	
Forward-Looking Model	-38,937.35	Z-statistic	$\frac{2,233.9}{25.94} = 86.13$
Knowledge Gradient	-41,171.25	p-value	0.000
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

During the estimation of the structural parameters, the variance-covariance matrix

$$\Sigma_0 = \begin{pmatrix} \begin{bmatrix} & \Sigma & \end{bmatrix} & \sigma_{z\nu} \\ \sigma'_{z\nu} & \sigma^2_{\nu} \end{pmatrix} \tag{27}$$

has a known component Σ, and an unknown border. We take the Cholesky decomposition of Σ_0, and parameterize its border with vector

$$\left[s_1, s_2, ..., s_6, \sqrt{1 - \sum_{j=1}^{6} s_j^2} \right]'$$

This yields the decomposed matrix in Table 15.

Table 15: Cholesky Decomposition of Bordered Variance-Covariance Matrix

$$C = \begin{pmatrix} 1. & -0.664415 & -0.305486 & -0.727454 & -0.0660357 & -0.159545 & s_1 \\ 0 & 0.747364 & 0.217517 & 0.29733 & 0.0471439 & 0.143938 & s_2 \\ 0 & 0 & 0.927019 & 0.128614 & 0.0370378 & 0.00187495 & s_3 \\ 0 & 0 & 0 & 0.604867 & 0.0352993 & 0.016755 & s_4 \\ 0 & 0 & 0 & 0 & 0.995389 & 0.00822777 & s_5 \\ 0 & 0 & 0 & 0 & 0 & 0.976461 & s_6 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sqrt{1-\sum_{j=1}^{6} s_j^2} \end{pmatrix}$$

The matrix $\overline{\Sigma}$, obtained by

$$\overline{\Sigma} = C'C \tag{28}$$

conforms to matrix Σ_0 in that the upper-left block is equal to Σ, and the element $\sqrt{1-\sum_{j=1}^{6} s_j^2}$ in C implies that resulting lower-right parameter σ_ν^2 is normalized to one. In addition, vector $\sigma_{z\nu}$ in Σ_0 is obtained by linear combinations of parameters $\{s_1..s_6\}$. Finally, at each iteration we check that matrix Σ is p.s.d., and penalize likelihood evaluations for guesses generating non-p.s.d. covariance matrices.

C State-Partitioning Decision Trees

C.1 Example

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 8 (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 9. Before estimation, we take draws from the joint distribution of characteristics (we use four draws here; this low number is convenient for illustration purposes). Suppose that we obtain draws $\{1, 2\}$, $\{1, 3\}$, $\{2, 2\}$, and $\{1, 3\}$, as represented in the top -left corner of Figure 9, 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 $x_2 = 2$.

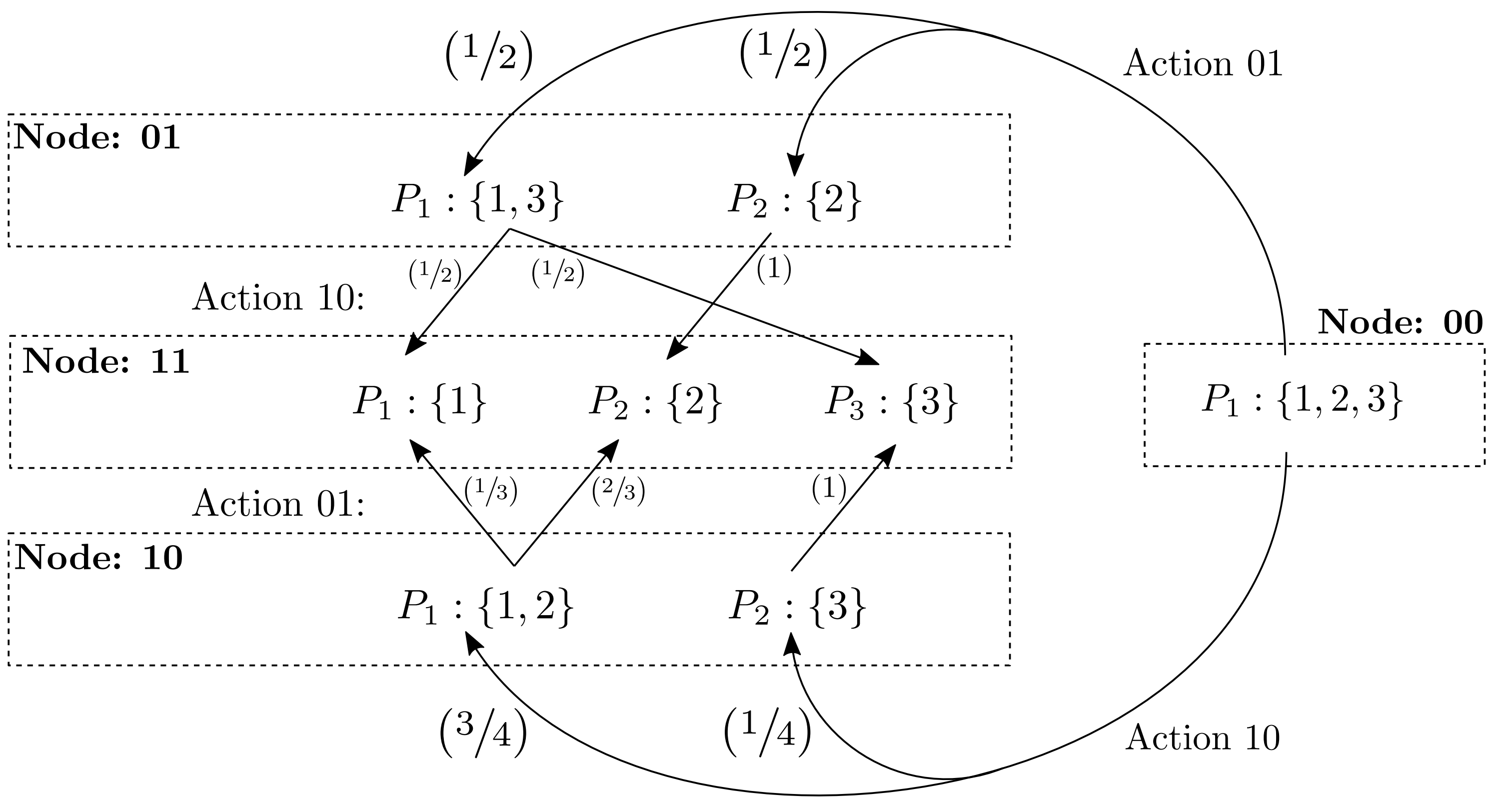
At node "00", any of the simulations may represent the true vehicle characteristics, but the decision maker does not know which one. Suppose the consumer decided to take search action 01 (top arrow), which informs her of characteristic x_2. She understands that she will transition from node 00 to node 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. Importantly, as the number of simulations increases, the draws represent the distribution of characteristics more closely, yielding estimator consistency.

Figure 9: Example of a State-Partitioning Decision Tree

Simulations

#	x_1	x_2	freq.
1)	1	2	1
2)	1	3	2
3)	2	2	1



C.2 Programming 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:

```
node{

    int id;
    vector of <partition> v_p;

}
```

Partition:

```
partition{

    vector of <action> v_a;
    vector of <int> simulations;
    vector of <terminal_utilities> v_tu;

}
```

Action:

```
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 9. 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 9, 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 memory pointers in coding 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.3 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 ν, 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.

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 ν 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.4 Unobserved Characteristic

Our goal is to maintain the relatively parsimonious tree representation described 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 16. Unless the cost of evaluating characteristic ν is prohibitively high, the DM should opt to immediately learn ν, because

Table 17: Example of Simulations with Continuous Signal $\widehat{\nu}$

Simulation #	x_1	x_2	$\widehat{\nu}$
1	1	1	1.5
2	1	2	0
3	2	2	1.5

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 16: Example of Simulations with Continuous Unknown Characteristic ν

Simulation #	x_1	x_2	ν
1	1	1	0.3
2	1	2	-1.3
3	2	2	0.6

An apparent solution would be to discretize the space of ν, 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 $\widehat{\nu} = \nu + \eta$, where η is normally distributed $N\left(0, \sigma_\eta^2\right)$. For estimation purposes, signal $\widehat{\nu}$ has a fixed support, based on the sparse grid proposed by Heiss and Winschel (2008). In this way, the number of support points for ν grows with the number of vehicles in the consumer's consideration set. Given the covariance matrix Σ, which includes the observable characteristics plus the unobserved characteristic ν, it is possible to write the distribution of $\widehat{\nu}$ conditional on the observable characteristics. For a specific parameter guess, we take draws of $\widehat{\nu}$, conditional on a simulation's observed characteristics, from the fixed support. The simulation probability of each support point $\widehat{\nu}_i$ is given by $\check{Pr}\left(\widehat{\nu}_i \middle| x_1, x_2\right) = \frac{f_{\widehat{\nu}}\left(\widehat{\nu}_i \middle| x_1, x_2\right)}{\sum_{j=1}^{NS} f_{\widehat{\nu}}\left(\widehat{\nu}_j \middle| x_1, x_2\right)}$ where $f_{\widehat{\nu}}\left(\widehat{\nu}_i \middle| x_1, x_2\right)$ is the density of the signal conditional on the observable characteristics, and NS is the number of sparse grid support points used. As NS grows, our simulation probability $\check{Pr}\left(\widehat{\nu}_i \middle| x_1, x_2\right)$ approaches the exact probability $Pr\left(\widehat{\nu}_i \middle| x_1, x_2\right)$ induced by the normal distribution.

Table 17 provides an example. Suppose that only three draws were taken, and the sparse grid of $\widehat{\nu}$ is $\{-1.5, 0, 1.5\}$, with simulation probabilities as defined above. According to Table 17, searching the unknown characteristic ν partitions the simulation set into subsets $\{1, 3\}$ and $\{2\}$. The expected value of ν can be calculated easily. Before any search activities take place, the DM is in an initial partition P_1, and the expected value of ν is calculated according to

$$
\begin{aligned}
E\left(\nu \mid P_1\right) = \frac{1}{3} & \left[E\left(\nu \mid x_1 = 1, x_2 = 1, \widehat{\nu} = 1.5\right)\right. \\
& + E\left(\nu \mid x_1 = 1, x_2 = 2, \widehat{\nu} = 0\right) \\
& \left. + E\left(\nu \mid x_1 = 2, x_2 = 2, \widehat{\nu} = 1.5\right)\right]
\end{aligned}
$$

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

$$
\begin{aligned}
E\left(\nu \mid P_2\right) = \frac{1}{2} & \left[E\left(\nu \mid x_1 = 1, x_2 = 1, \widehat{\nu} = 1.5\right)\right. \\
& \left. + E\left(\nu \mid x_1 = 2, x_2 = 2, \widehat{\nu} = 1.5\right)\right]
\end{aligned}
$$

The utility from the observed characteristics is calculated similarly.

As before, the simulation draws over $\widehat{\nu}$ 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 values.[14] 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.

C.5 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 17, 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. 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. 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 vehicles. Overall, our method ensures that as long as the number of simulations is appropriate, it yields a consistent estimator.

[14] Note also that the loss in efficiency from not updating the simulations of $\widehat{\nu}$ continuously with the covariance parameters decreases as the number of simulations increases.

C.6 Calculating the Likelihood of a Search Path

Once the decision-tree is created, according to the steps outlined in Section C.3, 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 18:

Table 18: Example of Simulations with Signal $\widehat{\nu}$

Simulation #	x_1	x_2	$\widehat{\nu}$
1	1	1	1.5
2	1	2	0
3	2	2	1.5
4	1	2	1.5

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 $\widehat{\nu}$ 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 could be incompatible with all of the draws. 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.

D Computational Details

All estimation took place on Amazon AWS hardware. The estimation of the main model was performed in parallel on "m4.10xlarge" instances. Estimation was programmed in C/C++.

References

Baye, M. R., J. Morgan, and P. Scholten (2004): "Price dispersion in the small and in the large: Evidence from an internet price comparison site," *The Journal of Industrial Economics*, 52(4), 463–496.

De los Santos, B., A. Hortaçsu, and M. R. Wildenbeest (2012): "Testing models of consumer search using data on web browsing and purchasing behavior," *American Economic Review*, 102(6), 2955–80.

Dinerstein, M., L. Einav, J. Levin, and N. Sundaresan (2018): "Consumer price search and platform design in internet commerce," *American Economic Review*, 108(7), 1820–59.

Einav, L., D. Knoepfle, J. Levin, and N. Sundaresan (2014): "Sales Taxes and Internet Commerce," *American Economic Review*, 104(1), 1–26.

Ellison, G., and S. F. Ellison (2009): "Search, obfuscation, and price elasticities on the internet," *Econometrica*, 77(2), 427–452.

Ellison, G., and A. Wolitzky (2012): "A search cost model of obfuscation," *The RAND Journal of Economics*, 43(3), 417–441.

Fradkin, A. (2019): "A Simulation Approach to Designing Digital Matching Platforms," *Working Paper*.

Frazier, P., W. Powell, and S. Dayanik (2009): "The knowledge-gradient policy for correlated normal beliefs," *INFORMS journal on Computing*, 21(4), 599–613.

Gabaix, X., and D. Laibson (2006): "Shrouded attributes, consumer myopia, and information suppression in competitive markets," *The Quarterly Journal of Economics*, 121(2), 505–540.

Heiss, F., and V. Winschel (2008): "Likelihood approximation by numerical integration on sparse grids," *journal of Econometrics*, 144(1), 62–80.

Hong, H., and M. Shum (2006): "Using price distributions to estimate search costs," *The RAND Journal of Economics*, 37(2), 257–275.

Hortaçsu, A., and C. Syverson (2004): "Product differentiation, search costs, and competition in the mutual fund industry: A case study of S&P 500 index funds," *The Quarterly Journal of Economics*, 119(2), 403–456.

Horton, J. J. (2014): "Misdirected Search Effort in a Matching Market: Causes, Consequences and a Partial Solution," in *Proceedings of the Fifteenth ACM Conference on Economics and Computation*, EC '14, pp. 357–357, New York, NY, USA. ACM.

Koulayev, S. (2014): "Search for differentiated products: identification and estimation," *The RAND Journal of Economics*, 45(3), 553–575.

Liang, A., X. Mu, and V. Syrgkanis (2017): "Optimal learning from multiple information sources," *Working Paper*.

Mann, H. B., and A. Wald (1943): "On stochastic limit and order relationships," *The Annals of Mathematical Statistics*, 14(3), 217–226.

McFadden, D. (1989): "A method of simulated moments for estimation of discrete response models without numerical integration," *Econometrica: Journal of the Econometric Society*, pp. 995–1026.

Powell, W. B. (2010): "The knowledge gradient for optimal learning," *Wiley Encyclopedia of Operations Research and Management Science*.

Sklar, A. (1959): "Fonctions de repartition an dimensions et leurs marges," *Publ. Inst. Statist. Univ. Paris*, 8, 229–231.

Vuong, Q. H. (1989): "Likelihood ratio tests for model selection and non-nested hypotheses," *Econometrica: Journal of the Econometric Society*, pp. 307–333.