A Method to Estimate Discrete Choice Models that is Robust to Consumer Search*

Jason Abaluck¶ Giovanni Compiani‡ Fan Zhang§

February 5, 2025

Abstract

We state conditions under which choice data suffices to identify preferences when consumers may not be fully informed about attributes of goods. Our approach can be used to test for full information, forecast how consumers will respond to information, and conduct welfare analysis when consumers are imperfectly informed. In a lab experiment, we successfully forecast the average response to new information when consumers engage in costly search. In data from Expedia, our method identifies which attribute was not immediately visible to consumers in search results, and we then use our model to compute the value of additional information.

*Thanks to Abi Adams, Judy Chevalier, Tim Christensen, Peter Hull, Joonhwi Joo, Magne Mogstad, Francesca Molinari, Barry Nalebuff, Aviv Nevo, Aniko Oery, Nicholas Ryan, Fiona Scott Morton, Brad Shapiro, Raluca Ursu, Miguel Villas-Boas, and seminar participants at Yale, SICS 2019, Marketing Science 2019, Caltech, QME 2019, Bologna, Northwestern, UT Austin, Bristol/Warwick, Chicago, Rice University, UNC, CEMFI, DICE, Penn, Cornell, Boston University, CREST, and Microsoft Research for helpful comments and suggestions. Tianyu Han, Jaewon Lee and Adam Sy provided excellent research assistance.

JEL codes: C5, C8, C9, D0, D6, D8.

¶Yale School of Management. Email: `jason.abaluck@yale.edu`.

‡University of Chicago, Booth School of Business. Email: `Giovanni.Compiani@chicagobooth.edu`.

§Nova School of Business and Eonomics. Email: `fan.zhang@novasbe.pt`.

1 Introduction

A large literature in marketing and economics documents that consumers are often imperfectly informed about the attributes of relevant products in ways that substantially alter their choices (see Honka, Hortaçsu, and Wildenbeest (2019) for a review). This pattern is found for relatively inexpensive items like groceries, as well as big-ticket and consequential purchases, such as cars, insurance and schooling (e.g., Bronnenberg and Vanhonacker (1996), Abaluck and Gruber (2011), Woodward and Hall (2012), Bronnenberg, Kim, and Mela (2016), Allcott, Lockwood, and Taubinsky (2019)). Given this, models that assume full information may generate wrong conclusions about welfare and—by construction—cannot be used to assess how consumers would respond to an information intervention.

In this paper, we state what we believe to be plausible sufficient conditions under which choice data alone suffices to recover preferences even if consumers are partially informed. In our baseline model, we assume that if consumers search, they do so in decreasing order of expected utility — a condition we make precise below. We show that if this condition is satisfied, along with a few additional restrictions we describe below, there is a function of choice probabilities which recovers preferences whether consumers are fully or partially informed. Our approach does not require the researcher to fully specify a structural search model beyond the expected utility assumption. Specifically, the expected utility assumption allows for a broad class of search models, including versions of classic sequential search (Weitzman 1979), the "directed cognition" model of Gabaix, Laibson, Moloche, and Weinberg (2006), or heuristic rules such as satisficing (Simon 1955) and simultaneously searching all goods with expected utility above a threshold.[1]

Recovering preferences under partial information has many applications. First, one can forecast the impact of informing consumers about attributes of goods prior to conducting such interventions, and compute the associated welfare benefits.[2] Second, our approach can inform firms' advertising strategies and the interface design of online platforms (e.g., by identifying product attributes that consumers care about but might not be currently aware of).[3] Third, in settings where one would otherwise assume full information, the expected utility assumption provides a generalization which allows for both full and partial information, thus permitting a more reliable normative evaluation of choices.[4] Finally, given preferences recovered by our approach, we show that it is possible to identify other primitives of

[1]The empirical literature suggests that canonical assumptions in all of these cases are often rejected by the data (Schwartz et al. (2002), Gabaix et al. (2006), Honka and Chintagunta (2017), Jindal and Aribarg (2018) in our four examples), limiting the applicability of fully specified search models.

[2]If an information intervention also reduces search costs, then the welfare gains from better choices can be viewed as a lower bound to the total increase in welfare.

[3]This is related to the literature on advertising content (e.g., Anderson, Ciliberto, and Liaukonyte (2013)).

[4]For example, when quantifying the gains from improving school quality (as measured by test scores), willingness to pay from actual choices may understate willingness to pay for parents who observe quality (Hastings and Weinstein 2008).

interest such as the distribution of search costs under a maintained model of search.

One can think of our approach as a data-driven method of isolating consumers who maximize utility. Consider the example of consumers shopping for laptops on an e-commerce platform. Some product information (e.g., price) is immediately *visible* on the results page, while other attributes (e.g., shipping speed) are relegated to the product page and require some time and effort to be uncovered; we refer to these attributes as *hidden.* Consumers may fail to maximize utility if they do not pay the cost to learn shipping speed for all products. Our *expected utility assumption* states that if you bother to check the shipping speed for a given laptop, you will first do so for any other laptop in the choice set that appears more promising to you based on the information immediately visible on the results page. This assumption implies that consumers who search the laptop with the best shipping speed *always choose the product that maximizes utility among all options* (which is not necessarily that with the best shipping speed). To see this, note that if some other laptop has higher utility than that with the best shipping speed, it must have higher expected utility and thus our assumption implies that it is searched and then chosen by the consumer. Further, only consumers who search the laptop with the best shipping speed are sensitive to shipping speed for that laptop. Therefore, by looking at the sensitivity of choices to the shipping speed of the laptop with the best shipping speed we are able to isolate consumers that behave *as if* they were fully informed; standard arguments for the full information case then recover their preferences.

Based on this argument, we show that a specific ratio of second derivatives of choice probabilities identifies preferences given partial information. Standard methods, based on ratios of first derivatives, could lead to attenuated estimates of preferences. In the above example, those approaches may erroneously conclude that consumers don't care much about shipping speed when, in fact, they are simply not aware that products differ along that dimension. Our second-derivative ratio recovers preferences even with partial information, and it *also* recovers preferences in the full information case. Thus, it provides robust estimation whether consumers are fully or partially informed. Further, we extend our model in several empirically relevant directions, including cases where the expected utility assumption is *not* satisfied. Specifically, we consider extensions where (i) search costs vary with observables (e.g., rank on a webpage), (ii) search reveals information that is unobservable to the researcher, (iii) either x or z is endogenous and valid instruments are available, and (iv) an outside option is available with utility known prior to search.

Our identification proof lends itself naturally to estimation and testing. If one can nonparametrically estimate choice probabilities as a function of product attributes, then our results can be used to directly recover preferences (Compiani 2022). We also suggest an alternative more parsimonious approach to estimate second derivatives that works well in simulations for larger numbers of goods. Further, one can use our result to test for full information by checking whether our "search-robust" estimates of preferences are equal to the conventional estimates based on first derivatives. This implies that one does not need to take an *a priori* stance on whether or not an attribute is uncovered only after searching a good. That hypothesis can be tested provided that the data contains at least one attribute that can be assumed to be known by consumers pre-search.

We conduct two data analyses to validate the usefulness of our approach assuming homogeneous preferences for product attributes. First, we show in a lab experiment involving e-book choice that we can recover full-information preferences over hidden discounts using only data on observed choices. Second, using data from hotel choice on Expedia, we show that our method correctly identifies as "hidden" the one attribute not immediately visible in search results (hotel location).

While our approach gains generality in relaxing the assumption that consumers are perfectly informed, it is comparatively data hungry, and it is potentially limited in the forms of heterogeneity it can accommodate. In particular, our main argument maintains that consumers have homogeneous preferences for product attributes and thus assumes that individuals who search the good with the highest value of the hidden attribute do not have systematically different preferences relative to the rest of the population. We show that identification does not rely completely on the homogeneous preferences assumption as our result can be extended to random coefficients models, provided we impose the (potentially strong) restriction of monotonicity – i.e., that everyone either likes or does not like the hidden attribute. Even with this identifying assumption, estimation of the random coefficients distributions requires parametric restrictions or estimation of derivatives of order higher than two, which may be infeasible in many applications given data limitations.

Estimation of our baseline model involves taking second-order derivatives of flexibly estimated functions as well as evaluating those functions at specific points in their domains, which could lead to slower-than-parametric convergence rates. Thus, we see it as most useful in settings with large data sets. Example applications might include: consumer product choice, evaluating whether shoppers are aware of per-unit prices when choosing among options of different sizes; school choice, evaluating whether parents are informed about average test scores or dropout rates for alternative schools; health plan choice, evaluating whether consumers are informed about plan network breadth; financial choice, evaluating whether consumers are informed about attributes of loans — such as the term length — that cannot be easily translated into dollar values; or housing choice, evaluating whether consumers are informed about measurable attributes, such as the noise level in a particular location. In all these cases, our model could also be used to test the impact of information interventions aimed at raising awareness of particular attributes. Of course, the limitations of our assumptions also apply: for example, we might think it plausible that all students prefer schools with lower dropout rates, but some students might also prefer schools with *lower* test scores due to tracking considerations; if so, our approach could give biased results for that variable.

Our result relates to several existing literatures. A theoretical and empirical literature models consumers as choosing from a possibly strict subset of the options available, their "consideration set."[5] Our model differs from much of the consideration set literature in that we consider the complementary problem of imperfect information at the level of attributes rather than goods. Specifically, in our setting, consumers are assumed to costlessly know some information for all goods, implying that, in general, consumer behavior will be affected by the visible attributes of all products, even those that are not searched.[6] A growing literature, including Mehta, Rajiv, and Srinivasan (2003), Kim, Albuquerque, and Bronnenberg (2010), Honka and Chintagunta (2017), Kim, Albuquerque, and Bronnenberg (2017), Ursu (2018) and Gardete and Hunter (2020), models consumers as searching products in order to uncover some of their attributes. We show that preferences can be recovered under our assumptions without committing to one specific structural search model.[7] This is similar in spirit to Compiani, Lewis, Peng, and Wang (2024), who study optimal rankings of products on e-commerce platforms; their model maintains the assumption of sequential search but is flexible as to what components of utility consumers learn via search. Another related literature studies whether consumers make informed choices by comparing the choices of regular consumers to that of a more informed subgroup. Bronnenberg, Dubé, Gentzkow, and Shapiro (2015) ask whether pharmacists make similar prescription drug choices to consumers, Handel and Kolstad (2015) ask whether better informed consumers make different health insurance choices, and Johnson and Rehavi (2016) study whether physicians treat differently when their patients are other physicians. Rather than identifying informed consumers, our paper develops a data-driven way of identifying consumers who maximize utility (despite not necessarily searching all goods) and whose choices can thus be used to recover preferences.

The rest of the paper is organized as follows. Section 2 lays out our formal framework and proves our key identification results, Section 3 considers several empirically important extensions such as endogenous attributes, Section 4 discusses the model assumptions and their testability, and the (coun-

[5]Roberts and Lattin (1991), Goeree (2008), Conlon and Mortimer (2013) and Gaynor, Propper, and Seiler (2016) — among others — estimate preferences when consumers may only consider some alternatives. Manzini and Mariotti (2014) establish that one can recover consideration probabilities as well as preferences if the data contains choices from every possible subset of the feasible set of goods. Abaluck and Adams (2017) show that identification can be achieved even without this type of variation under certain models of consideration set formation. Cattaneo, Ma, Masatlioglu, and Suleymanov (2020) and Barseghyan, Coughlin, Molinari, and Teitelbaum (2021) study partial identification of a general model with heterogeneous consideration sets. Agarwal and Somaini (2022) propose a model in which consumers choose from choice sets that are unobserved to the researcher due to information frictions or supply-side rationing, and show how instrumental variables enable identification of preferences as well as the rule determining which products are included in the choice set.

[6]The recent theoretical literature on this question includes Branco, Sun, and Villas-Boas (2012), Ke, Shen, and Villas-Boas (2016) and Gabaix (2019).

[7]In this sense, our paper is also related to the literature on "attribute non-attendance." There is one special case where the problem of imperfect information about attributes has been often addressed in the existing literature. This is the case in which all attributes can be expressed in dollar terms. For example, consumers should not care whether a health insurance plan saves them \$100 in premiums or out of pocket costs (see Abaluck and Gruber (2011)), or whether a light bulb saves them money in upfront costs or shelf life (as in Allcott and Taubinsky (2015)). If one dollar-equivalent attribute is assumed to be visible to consumers, it can provide a benchmark for how consumers should respond to a hidden dollar-equivalent attribute. However, in many cases, attributes cannot easily be translated into dollars without first estimating consumer preferences. In these cases, our results still allow one to recover preferences given imperfectly informed consumers.

terfactual) questions that can be addressed using our approach, Section 5 provides details of estimation, Section 6 offers a guide to practitioners, Sections 7 and 8 report results from our experiment and Expedia application, respectively, and Section 9 concludes.

2 Model and Identification

2.1 Setup

There are $J \geq 2$ goods indexed by $j = 1, ..., J$ with attributes x_j observed by consumers and the econometrician and an attribute z_j observed by the econometrician but not necessarily by consumers.[8] In order to focus on the intuition underlying our key results, we make five simplifying assumptions in our exposition here. First, we let x_j be scalar for all j; our results immediately extend to the case of vector-valued x_j's at the cost of some extra notation. Second, we focus on the case where z_j is also a scalar; we consider the case with multivariate z_j in Appendix A.3. Third, we assume that x_j and z_j are continuously distributed (Appendix A.8 shows that an analogous argument applies to the case with discrete attributes provided that at least one attribute is continuous). Fourth, we assume that the utility that individual i derives from good j is linear in x_j, z_j and an idiosyncratic shock ϵ_{ij} that is observed by consumers prior to search (we consider in Appendix A.6 the case where unobservables are seen only after search). Fifth, we focus on the case where both x_j and z_j are exogenous (Section 3.1 discusses how to deal with endogeneity). We formalize these assumptions as follows.

Assumption 1. The utility that individual i derives from good j is $U_{ij} = \alpha x_j + \beta z_j + \epsilon_{ij}$, where x_j, z_j are continuous scalars, and $\mathbf{x} = (x_1, \ldots, x_J)$ and $\mathbf{z} = (z_1, \ldots, z_J)$ are independent of $\epsilon_{\mathbf{i}} = (\epsilon_{i1}, \ldots, \epsilon_{iJ})$. The consumer observes x_j, ϵ_{ij} for all j prior to search, but needs to search good j to uncover z_j.

We assume that consumers form expectations on z_j based on x_j: $E(z_j|x_j) = \gamma_0 + \gamma_1 x_j$. Then, the expected utility from good j before engaging in search takes the form $EU_{ij} = \beta\gamma_0 + (\alpha + \beta\gamma_1)x_j + \epsilon_{ij}$. This nests the case where consumers have (linear) rational expectations as well as the case where they do not update at all about z_j based on observed attributes of goods ($\gamma_1 = 0$). We let $\tilde{z}_j \equiv z_j - (\gamma_0 + \gamma_1 x_j)$ be the *surprise utility* revealed by searching j relative to consumers' expectations.

Next, we state the assumptions that characterize the class of search models we consider.

Assumption 2. (i) Consumer i searches goods in decreasing order of EU_{ij}.

(ii) Conditional on having utility $\bar{u}$ in hand (i.e., having uncovered a maximum utility $\bar{u}$ in the search process so far), consumer i searches j if and only if $g_i(x_j, \epsilon_{ij}, \bar{u}) \geq 0$ where g_i is a decreasing

[8] Our model also permits the more general case where attributes are potentially both good and individual-specific, but we write x_j and z_j rather than x_{ij} and z_{ij} for notational simplicity.

function of $\bar{u}$.[9]

(iii) Consumers choose the good which maximizes utility among searched goods.

(iv) Only the value of z_j is unknown to consumers prior to search, and search fully reveals z_j.

We discuss these conditions at length in Section 4.1. To briefly clarify, Assumption 2(i) states that consumers search goods in descending order of expected utility prior to search. Assumption 2(ii) states that consumers decide whether or not to search a good based on their utility in hand and the the information they have available about the good they are considering searching — i.e., (x_j, ϵ_{ij}). This rules out, for example, a sequential search protocol whereby one stops searching after discovering a good with large z irrespective of utility in hand.[10] We subscript the function g by i to emphasize that the function may depend on any individual (unobserved) heterogeneity in utility or search. For example, in a Weitzman search model, the stopping rule would depend on consumer i's reservation value, which in turn depends on i's search cost. Assumption 2(iii) states that consumers must search a good before choosing it. Assumption 2(iv) states that the econometrician observes all the information which is revealed by search, and that search is fully informative about the hidden attribute.

We pause here to highlight that Assumption 2 accommodates several commonly used models of search.

Example 1 (Sequential Search)**.** *Suppose that consumers search sequentially and consumer i must pay a cost c_i every time she uncovers the z attribute for a good. Further, assume that the consumer believes that $\tilde{z}$ is distributed according to the prior $F_{\tilde{z}}$ i.i.d. across goods. Then, following Weitzman (1979), the consumer will rank goods according to their reservation value rv'_{ij} defined implicitly by*

$$c_i = \int_{rv'_{ij}}^{\infty} \left(u - rv'_{ij}\right) dF_{U_{ij}|x_j}(u) = \int_{rv_i}^{\infty} \beta \left(t - rv_i\right) dF_{\tilde{z}}\left(t\right) \tag{1}$$

where $rv_i \equiv \frac{rv'_{ij} - \beta\gamma_0 - (\alpha+\beta\gamma_1)x_j - \epsilon_{ij}}{\beta}$ and the last step follows from a change of variable. We can interpret rv_i as the reservation value in units of $\tilde{z}$. To see this, note that consumer i searches goods in descending order of $\beta\gamma_0 + (\alpha + \beta\gamma_1)x_j + \beta rv_i + \epsilon_{ij}$ (or, equivalently, of $EU_{ij} = \beta\gamma_0 + (\alpha + \beta\gamma_1)x_j + \epsilon_{ij}$) and for each good j', she chooses to uncover $z_{j'}$ if and only if the maximum utility secured so far is lower than $\beta\gamma_0 + (\alpha + \beta\gamma_1)x_{j'} + \beta rv_i + \epsilon_{ij'}$. Once she stops searching, she maximizes utility among the searched goods. Thus, Assumption 2 is satisfied with $g_i\left(x_j, \epsilon_{ij}, \bar{u}\right) = \beta\gamma_0 + (\alpha + \beta\gamma_1)x_j + \beta rv_i + \epsilon_{ij} - \bar{u}$.

Example 2 (Directed Cognition Model)**.** *As in the model of Gabaix, Laibson, Moloche, and Weinberg (2006), suppose that consumers rank goods in terms of expected utility and myopically check whether*

[9] Assumption 2(ii) can be weakened to allow the function g_i to depend on a good-specific unobservable, such as search costs; however, good-specific search costs may lead to violations of Assumption 2(i). In Section 3.2, we extend our model to permit search costs to vary across goods with observable factors.

[10] Assumption 2(ii) does accommodate simultaneous search models in which consumers decide which goods to uncover based on expected utilities and then proceed to jointly search them. In this case, consumers don't sequentially uncover utilities and the function g_i does not vary with its second argument.

searching the next good is worth the cost. The directed cognition model has the same search order as the Weitzman model but a different g_i function; consumers myopically check whether $g_i(x_j, \epsilon_{ij}, \bar{u}) = E_{\tilde{z}} \max(U_{ij} - \bar{u}, 0) - c_i \geq 0$ to decide whether to search good j.

Example 3 (Satisficing)**.** *Suppose that consumer i searches in order of expected utility and stops whenever utility in hand is above a threshold τ_i. Then, Assumption 2 is satisfied with $g_i(x_j, \epsilon_{ij}, \bar{u}) = \tau_i - \bar{u}$.*

Example 4 (Full Information)**.** *The full information model is subsumed within the previous example by letting $\tau_i = \infty$ for all i.*

Example 5 (Simultaneous Search)**.** *Suppose that consumer i simultaneously searches all goods that have expected utility above a threshold $\tilde{\tau}_i$. Then, Assumption 2 is satisfied with $g_i(x_j, \epsilon_{ij}, \bar{u}) = \beta\gamma_0 + (\alpha + \beta\gamma_1)x_j + \epsilon_{ij} - \tilde{\tau}_i$.*

Our results will not require the researcher to take a stand on the specific model of search that consumers follow provided that our assumptions are met. Therefore, as illustrated by the examples above, the approach will be agnostic as to whether consumers search sequentially or simultaneously, are forward-looking or myopic and have biased or unbiased beliefs, among other things. In contrast, fully specifying a structural model requires one to take a stance on each of these dimensions.

2.2 Identification

We assume throughout without loss that $\beta > 0$, i.e. we treat z_j as an attribute that consumers value in good j.[11] Further, we let product 1 be the good with the highest value of $\tilde{z}$. In other words, good 1 delivers the largest positive surprise relative to consumers' expectations. Note that $\tilde{z}_j$ collapses to z_j when $\gamma_1 = 0$, i.e. when consumers do not form any expectations on z (so in the case of myopic expectations, good 1 simply has the highest value of z). Here, we consider the case where the parameter γ_1 governing the way in which consumers form expectations is known to the researcher. For instance, if consumers have rational expectations, γ_1 can simply be estimated by regressing z_j on x_j. Knowledge of γ_1 implies knowledge of $\tilde{z}_j$ for all j, which means that the researcher knows the identity of good 1 in any given choice set. Section 3.5 considers the case where γ_1 is unknown to the researcher.

We are now ready to state and prove a key lemma.

Lemma 1. *Let Assumptions 1 and 2 hold. If consumer i searches good 1 (i.e., the good with the highest value of $\tilde{z}$), then i chooses the utility-maximizing good.*

[11] This is without loss, since Assumption 2 implies that an increase in U_{ij} can only induce consumer i to switch from not choosing j to choosing j, but never vice versa. (see footnote 32 for a formal argument). Thus, by the chain rule, the sign of β is identified by the sign of $\frac{\partial s_j}{\partial z_j}$, where s_j is the choice probability function for good j from the data.

Proof. If good 1 is searched but utility is not maximized, then for some unsearched j, $U_{ij} > U_{i1}$. Since $\tilde{z}_1 \geq \tilde{z}_j$, it must be that $EU_{ij} > EU_{i1}$. But by Assumption 2(*i*), this implies that good j is searched, which is a contradiction. $\square$

Note that Lemma 1 does *not* imply that good 1 always maximizes utility if it is searched. Rather, it implies that if good 1 is searched, the utility-maximizing good will also be searched (whether it is good 1 or not) and thus the consumer will choose that good. The lemma also does *not* mean that consumers searching good 1 are fully informed (in a search model they typically will not be), but just that those consumers act *as if* they were fully informed.

Lemma 1 will have far-reaching implications. To understand why, it will be convenient to define the choice probability for good j as:

$$s_j \equiv P\left(\left\{U_{ij} = \max_k U_{ik} \text{ for } k \in \mathcal{G}_i\right\} \cap \{j \in \mathcal{G}_i\}\right) \tag{2}$$

where $\mathcal{G}_i$ denotes the set of searched goods for individual i. Note that this probability is computed by integrating over any individual-specific unobserved heterogeneity in utility or search, while fixing the choice set attributes $\mathbf{x} \equiv [x_1, \ldots, x_J]$ and $\mathbf{z} \equiv [z_1, \ldots, z_J]$. Therefore, s_j is a function of $(\mathbf{x}, \mathbf{z})$, but we will often omit the dependence from the notation. Throughout the paper, the sources of individual-level unobserved heterogeneity will vary with the specific models we consider, so the symbol P will denote integrals over different distributions depending on the context.[12]

Now, Lemma 1 implies that z_1 only impacts choice probabilities for individuals who maximize utility. Therefore, looking at $\frac{\partial s_1}{\partial z_1}$ will isolate individuals who maximize utility and allow us to recover preferences using standard arguments. To formalize this, note that Lemma 1 implies that we can write:

$$s_1 = P\left(U_{i1} \geq U_{ik}\ \forall k\right) - P\left(\{U_{i1} \geq U_{ik}\ \forall k\} \cap \{\text{for some } j \neq 1,\ EU_{ij} \geq EU_{i1} \text{ and } g_i(x_1, \epsilon_{i1}, U_{ij}) \leq 0\}\right) \tag{3}$$

In other words, the probability that good 1 is chosen is the probability that good 1 is utility-maximizing minus the probability of the only type of mistakes that consumers searching good 1 can make, i.e. failing to search good 1 *even though it is utility-maximizing*. Failing to search good 1 requires that there exists some other good j with $EU_{ij} \geq EU_{i1}$ and utility high enough that $g_i(x_1, \epsilon_{i1}, U_{ij}) \leq 0$. The other type of mistake, i.e. choosing good 1 when it is not utility-maximizing, is ruled out by Lemma 1 and thus does not feature in (3).

We will now use equation (3) to show our key result, i.e. that the preference parameters α and β are identified from the second derivatives of function s_1. Since the distribution of ϵ is unrestricted, we impose the following normalizations without loss: $\alpha = 1$ (scale) and $\epsilon_{i\tilde{j}} = 0$ for some $\tilde{j}$ and all i (location).

[12] For instance, in the Weitzman model of Example 1, the probability is taken over the joint distribution of the preference shocks ϵ_i and the search costs c_i.

Lemma 2. *Let Assumptions 1 and 2 hold. Further, assume that s_1 is twice differentiable and that $\frac{\partial^2 s_1}{\partial z_1 \partial x_j}(\mathbf{x}^*, \mathbf{z}^*) \neq 0$ for some $(\mathbf{x}^*, \mathbf{z}^*)$ and some $j \neq 1$. Then,*

$$\frac{\partial^2 s_1}{\partial z_1 \partial z_j}(\mathbf{x}^*, \mathbf{z}^*) \Big/ \frac{\partial^2 s_1}{\partial z_1 \partial x_j}(\mathbf{x}^*, \mathbf{z}^*) = \frac{\beta}{\alpha} = \beta, \tag{4}$$

so that β is identified. In addition, the distribution of ϵ is nonparametrically identified if supp $\left(\epsilon_k - \epsilon_{\tilde{j}}\right)_{k \neq \tilde{j}}$ $\subset \left\{(\alpha + \beta\gamma_1)\left(x_{\tilde{j}} - x_k\right)_{k \neq \tilde{j}} : \gamma_1 (x_{\tilde{j}} - x_k)_{k \neq \tilde{j}} = \left(z_{\tilde{j}} - z_k\right)_{k \neq \tilde{j}} \text{ for some } (\mathbf{x}, \mathbf{z}) \text{ in its support}\right\}$ and the support of $(\tilde{z}_1, \ldots, \tilde{z}_J)$ contains a point such that $\tilde{z}_j = \tilde{z}_k$ for all j, k.

Proof. To facilitate intuition, we focus on the case with $J = 2$ goods here. Appendix A.1 contains the proof for the general case with $J \geq 2$ goods. First, we prove equation (4), often suppressing the subscript i in what follows. The probability of choosing good 1 can be written as:

$$s_1 = P(U_1 > U_2) - P\left(\{U_1 > U_2\} \cap \{\mathcal{G} = \{2\}\}\right) \tag{5}$$

where again $\mathcal{G}$ denotes the set of searched goods. This follows because (i) if good 1 is utility-maximizing, consumers will always choose it unless they search only good 2; and (ii) the other type of mistake (choosing good 1 when it is not utility-maximizing) is ruled out by Lemma 1.

Let $\tilde{u}_j \equiv \alpha x_j + \beta z_j$ be the mean utility of good j, so that $U_{ij} = \tilde{u}_j + \epsilon_{ij}$, and let $(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^*, \mathbf{z}^*)$. Our goal will be to show that both z_2 and x_2 only impact $\frac{\partial s_1}{\partial z_1}$ via $\tilde{u}_2$. This, in turn, implies that $\frac{\partial^2 s_1}{\partial z_1 \partial z_2} = \frac{\partial^2 s_1}{\partial z_1 \partial \tilde{u}_2} \frac{\partial \tilde{u}_2}{\partial z_2}$ and $\frac{\partial^2 s_1}{\partial z_1 \partial x_2} = \frac{\partial^2 s_1}{\partial z_1 \partial \tilde{u}_2} \frac{\partial \tilde{u}_2}{\partial x_2}$, implying the result in equation (4). To this end, note that:

$$\begin{aligned} P\left(\{U_1 > U_2\} \cap \{\mathcal{G} = \{2\}\}\right) = \\ P(\{U_1 > U_2\} \cap \{EU_2 > EU_1\} \cap \{g(x_1, \epsilon_1, U_2) \leq 0\}) = \\ P(\{U_1 > U_2\} \cap \{g(x_1, \epsilon_1, U_2) \leq 0\}) - P(\{EU_1 > EU_2\} \cap \{g(x_1, \epsilon_1, U_2) \leq 0\}) \end{aligned} \tag{6}$$

where the second line follows since $EU_1 > EU_2$ implies $U_1 > U_2$ and thus $P\left(\{EU_1 > EU_2\} \cap \{g(x_1, \epsilon_1, U_2) \leq 0\}\right) = P\left(\{U_1 > U_2\} \cap \{EU_1 > EU_2\} \cap \{g(x_1, \epsilon_1, U_2) \leq 0\}\right)$. The second term on the last line of display (6) is not a function of z_1. The first term is only a function of x_2 and z_2 via $\tilde{u}_2$. This, along with equation (5), implies that both z_2 and x_2 only impact $\frac{\partial s_1}{\partial z_1} = \frac{\partial P(U_1 > U_2)}{\partial z_1} - \frac{\partial P(\{U_1 > U_2\} \cap \{\mathcal{G} = \{2\}\})}{\partial z_1}$ via $\tilde{u}_2$. This, together with the normalization $\alpha = 1$, proves equation (4).

We now show that the distribution of ϵ can be identified. Consider choice sets where $\tilde{z}_j$ is the same for all j. In this case, the good with the highest expected utility also has the highest realized utility. Thus, consumers who search in descending order of expected utility always choose the utility-maximizing good and we can write:

$$s_{\tilde{j}} = P(\epsilon_k - \epsilon_{\tilde{j}} \leq (\alpha + \beta\gamma_1)(x_{\tilde{j}} - x_k) \quad \forall k). \tag{7}$$

This is a standard full-information demand model and we can follow the argument in Section 4.2 of Berry and Haile (2014) to identify the joint distribution of ϵ. Specifically, letting $\mathbf{x}$ vary — and relying on the fact that the coefficient $\alpha + \beta\gamma_1$ is identified given knowledge of γ_1 — one can use (7) to trace out the distribution of $\left(\epsilon_k - \epsilon_{\tilde{j}}\right)_{k \neq \tilde{j}}$ over the set in the statement of the theorem. This, paired with the location normalization $\epsilon_{i\tilde{j}} = 0$, gives identification of the distribution of ϵ. □

This result shows that the preferences parameter β is identified from our ratio of second derivatives and that the entire distribution of ϵ can be traced out nonparametrically under additional support restrictions. Specifically, this second argument relies on varying $\mathbf{x}$ and $\mathbf{z}$ in such a way that $\tilde{z}_j = \tilde{z}_k$ for all j, k, which is equivalent to $\gamma_1(x_j - x_k) = z_j - z_k$. Thus, we require the support of ϵ to be contained in the set of $(\alpha + \beta\gamma_1)\,\mathbf{x}$ values that satisfy this condition (for some value of $\mathbf{z}$). Note that in the case where $\gamma_1 = 0$, the assumption reduces to requiring that the support of ϵ be contained in the support of $\mathbf{x}$, which is in line with restrictions used in the literature (Berry and Haile 2014). In practice, nonparametrically estimating the distribution of ϵ may be challenging with limited data. Indeed, in our empirical applications, we leverage standard parametric assumptions.

Finally, we show that in many models of interest, identifying preferences based on the ratio of *first* derivatives leads to understating consumers' taste for z. For this result, we make two mild additional assumptions: (i) that the function $g_i(x_j, \epsilon_{ij}, \bar{u})$ is weakly increasing in x_j, and (ii) that $\alpha + \beta\gamma_1 \geq 0$. Condition (i) is satisfied in all the search model considered above (Examples 1–5) when the coefficient on x in utility is positive and corresponds to the mild requirement that consumers are (weakly) more prone to searching a good the higher the value of x for that good. Condition (ii) is also intuitive and it requires that the effect of x_j through expectations about z_j (i.e., $\beta\gamma_1$) not flip the sign of the overall effect. For example, if x_j captures how inexpensive the product is and z_j is its quality, then it must be that consumers still value a decrease in price even after accounting for the fact that they expect lower quality.

Lemma 3. *Let Assumptions 1 and 2 hold. Further, assume that $g_i(x_j, \epsilon_{ij}, \bar{u})$ is weakly increasing in x_j and that $\alpha + \beta\gamma_1 \geq 0$. Then, for all j, k,*

$$\left| \frac{\partial s_j}{\partial z_k} \Big/ \frac{\partial s_j}{\partial x_k} \right| \leq |\beta|. \tag{8}$$

Proof. See Appendix A.2. □

The results in Lemmas 2 and 3 suggest a natural approach to test for full information. Under the null hypothesis of full information, $s_j = P\left(U_{ij} \geq U_{ik}\ \forall k\right)$ and therefore, by the chain rule,

$$\frac{\partial s_j}{\partial z_k} \Big/ \frac{\partial s_j}{\partial x_k} = \frac{\partial^2 s_1}{\partial z_1 z_{j'}} \Big/ \frac{\partial^2 s_1}{\partial z_1 \partial x_{j'}} = \beta \tag{9}$$

for all j, k and all $j' \neq 1$. On the contrary, when consumers are unaware of z for some goods, then the ratios of first derivatives need not be equal to the ratios of the second derivatives. In particular, Lemma 3 showed that the ratios of first derivatives will tend to be smaller in magnitude within the class of search models we consider. This motivates the testing approach in the following lemma.

Lemma 4. *Under Assumption 1, if consumers are fully informed, then* $\left|\frac{\partial s_j}{\partial z_k}\Big/\frac{\partial s_j}{\partial x_k}\right| = \left|\frac{\partial^2 s_1}{\partial z_1 \partial z_{j'}}\Big/\frac{\partial^2 s_1}{\partial z_1 \partial x_{j'}}\right|$ *for all* j, k *and all* $j' \neq 1$*. Under Assumptions 1 and 2,* $\left|\frac{\partial s_j}{\partial z_k}\Big/\frac{\partial s_j}{\partial x_k}\right| \leq \left|\frac{\partial^2 s_1}{\partial z_1 \partial z_{j'}}\Big/\frac{\partial^2 s_1}{\partial z_1 \partial x_{j'}}\right|$ *for all* j, k *and all* $j \neq 1$*. Thus, it is possible to test the hypothesis of full information by checking whether the ratios of first derivatives are attenuated relative to the ratios of second derivatives.*

A natural question is whether the testable implications in Lemma 4 are still valid when Assumption 1 is violated and, in particular, when the coefficients on product attributes are heterogeneous across consumers. One might worry that when consumers have heterogeneous preferences, the ratios of first derivatives might be attenuated relative to the ratios of second derivatives even under the null hypothesis of full information. In Appendix C, we provide verifiable sufficient conditions that rule this out when the ϵ shocks are Gumbel distributed and the coefficients can be signed (e.g., β is distributed on the positive reals); the latter is needed to ensure that the identity of good 1 is consistent across consumers. Under these conditions, the testing approach based on second derivatives is valid even in the presence of heterogeneity in α or β. Further, we show that this is not generally the case when $\beta = 0$ for a positive mass of consumers. Intuitively, if some consumers don't care about z at all, this will tend to lead to attenuation of the first-derivative ratios even under full information. However, we show that using a slightly different approach allows us to distinguish between this scenario and the case where consumers search (and have homogeneous β). More specifically, we use the fact that ratios of derivatives of the same order (i.e., $\frac{\partial^2 s_1/\partial z_1 \partial z_2}{\partial^2 s_1/\partial z_1 \partial x_2}$ and $\frac{\partial^2 s_2/\partial z_2 \partial z_1}{\partial^2 s_2/\partial z_2 \partial x_1}$) are equal under the assumption of full information, even when $\beta = 0$ for some consumers. In contrast, these ratios will in general not be the same when consumers engage in search. This discrepancy allows us to empirically distinguish the two cases.

So far we have discussed how to adjust our testing approach to accommodate heterogeneity in β on the positive or negative reals. Identifying the distribution of the random coefficients in our setting is more challenging and in general will require taking higher-order derivatives or maintaining parametric assumptions on the distribution of the random coefficients. We provide some results on this in Appendix A.10.

3 Extensions

We now consider a few extensions to our baseline model that are relevant for empirical work.

3.1 Endogenous attributes

So far, we have assumed that the observed product attributes are independent of all unobservables. This can be restrictive, especially in settings in which product attributes — notably price — are chosen by firms who might know more about preferences or product attributes than is captured by the observed data. As highlighted by a large literature (e.g., Berry, Levinsohn, and Pakes (1995)), this typically leads to correlation between the attributes chosen by firms and product-level unobservables.

Here we consider an extension of our model that allows for endogenous product attributes. We specify the utility that consumer i gets from good j as

$$U_{ij} = \alpha x_j + \beta z_j + \lambda p_j + \xi_j + \epsilon_{ij} \tag{10}$$

where p_j denotes the endogenous characteristic and ξ_j is a product-specific characteristic that is known by consumers before search, but is not observed by the researcher. We consider both the case where p_j is visible to consumers without search and that in which consumers need to search good j to uncover p_j (in which case, with a slight abuse of notation, p_j coincides with z_j). If p_j is price, the first scenario corresponds to settings such as e-commerce where typically price is visible on the results page and does not require any further clicking by the user. On the other hand, the second scenario covers cases in which price is itself the object of consumer search (there is a large literature on this, particularly in relation to the often observed price dispersion for relatively homogeneous goods; see, e.g., Stahl (1989), Hong and Shum (2006) and Hortaçsu and Syverson (2004)). We show identification of preferences for each of these two cases. To this end, we introduce two mutually exclusive variants of assumptions 2(i)-2(ii).

Assumption 3. (i) The attribute p_j is visible to consumers prior to search and consumers form expectations on z_j using $E(z_j|x_j, p_j) = \gamma_0 + \gamma_1 x_j + \gamma_{1,p} p_j$. Further, consumers search in descending order of expected utility $EU_{ij} = \delta_j + \epsilon_{ij}$, where $\delta_j \equiv \beta\gamma_0 + (\alpha + \beta\gamma_1)x_j + (\lambda + \beta\gamma_{1,p})p_j + \xi_j$. Conditional on having utility $\bar{u}$ in hand, consumer i searches j if and only if $g_i(\delta_j, \epsilon_{ij}, \bar{u}) \geq 0$ where g_i is increasing in δ_j and decreasing in $\bar{u}$.

(ii) The attribute p_j is uncovered by consumers only upon searching good j. In this case, z_j coincides with p_j and consumers form expectations on p_j using $E(p_j|x_j) = \gamma_0 + \gamma_1 x_j$. Further, consumers search in descending order of expected utility $EU_{ij} = \underline{\delta}_j + \epsilon_{ij}$, where $\underline{\delta}_j \equiv \beta\gamma_0 + (\alpha + \beta\gamma_1)x_j + \xi_j$. Conditional on having utility $\bar{u}$ in hand, consumer i searches j if and only if $g_i(\underline{\delta}_j, \epsilon_{ij}, \bar{u}) \geq 0$ where g_i is increasing in $\underline{\delta}_j$ and decreasing in $\bar{u}$.

Like Assumptions 2(i) and 2(ii), Assumption 3 states that consumers search in descending order of expected utility and decide whether to search good j based on the utility in hand and the information they have on good j prior to search. In Appendix A.4, we show that our model falls within the class of demand systems studied by Berry and Haile (2014). Thus, we can invoke their results to establish nonparametric identification of the structural choice probability functions — i.e., the functions that map all market-level attributes, including the unobservables ξ, into the probability of purchase. Of course, this requires valid instruments for the endogenous variables, but the requirements are no more demanding than in the case of full information. In particular, standard arguments can be applied to motivate the usual instruments for price, such as cost shifters, exogenous attributes of competing products, and prices of the same products in other markets. One caveat is that, to ensure that the exclusion restriction is satisfied, it's necessary to assume that consumers do not know these instruments or, if they do, that they do not use them to form expectations on z. For instance, one might assume that consumers are not aware of production cost shifters when making their choices. Once the structural choice probability functions are identified, one may apply our results in Section 2.2 to identify the preference parameters in (10). Note that, as in the case with full information, recovering the structural probability functions when some of the product attributes are endogenous requires variation across many markets. This is more demanding of the data relative to the case without endogeneity where just a few markets suffice in principle (see Section 4.2 for more discussion of this point).

3.2 Allowing for variables affecting search but not utility

One important case in which the assumption that consumers search in descending order of expected utility (Assumption 2(i)) is likely to fail is when some factors impact search costs but not utility. An example might be search position for online purchases. Arguably, search position impacts the order in which people search but often has no direct impact on utility conditional on search (Ursu 2018). In this case, consumers might first search items with higher search positions even if they do not have higher expected utility. For example, if we randomly assign search order, this is likely to impact choices even though we are not changing the utility of each item conditional on search. Another example is advertising, which entices consumers to search advertised goods but may not affect their utility.

Our model can be extended to deal with cases where the factors impacting search but not utility are observable and the sign of their impact on search probabilities is known (such as position in a list of e-commerce results). Denoting this variable by r_j, suppose that r_j is observed by the researcher and that higher values of r_j make a good weakly more likely to be searched. Now, rather than assuming that goods are searched based on EU_{ij} alone, we assume that goods are searched based on $m(EU_{ij}, r_j)$, where m is strictly increasing in both EU_{ij} and r_j. We show in Appendix A.5 that our identification argument from Lemma 2 continues to hold under additional assumptions, including an IIA restriction. In spite of these additional, stronger assumptions, we show that our results are not sensitive to implementing this alternative model in our empirical application in Section 8.

3.3 Unobservables revealed by search

Our baseline model focused on the case where the attribute(s) z revealed by searching a good are entirely observed by the researcher. However, it is easy to imagine settings in which the data does not capture all of the information that consumers acquire through search. Indeed, the existing literature often models search as the process whereby the idiosyncratic preference shocks — ϵ_{ij} in our notation — are revealed (e.g., Kim, Albuquerque, and Bronnenberg (2010), Ursu (2018), Moraga-González, Sándor, and Wildenbeest (2021)). To accommodate this, we consider a modification of our model where the shock ϵ_{ij} only becomes known to consumer i upon searching good j (along with z_j). In other words, consumers know x_j for all j prior to search and decide whether to acquire ϵ_{ik} and z_k for any given good k through search. As a result, the expected utility in Assumption 2(i) now takes the form $EU_j = \beta\gamma_0 + (\alpha + \beta\gamma_1)x_j$, the rule determining whether a good is searched in Assumption 2(ii) is now $g_i(x_j, \overline{u}) \geq 0$, and Assumption 2$(iv)$ is dropped.

In this case, the hidden part of utility is not fully observed by the researcher and thus Lemma 1 does not hold (in fact, it is not even possible to identify good 1). However, if consumers tend to value the x attribute (i.e., $\alpha + \beta\gamma_1 \geq 0$), they will search in descending order of it, implying they will always search the good with the highest x. Using this fact, we show in Appendix A.6 that the ratio of second derivatives $\frac{\partial^2 s_j}{\partial z_j \partial z_k} \Big/ \frac{\partial^2 s_j}{\partial z_j \partial x_k}$ recovers $\frac{\beta}{\alpha}$ provided that one chooses good k to be the good with the highest value of x, which could coincide with good 1 (here, j need not be the good with the highest value of $\tilde{z}$). Recall that in our baseline model, $\frac{\partial^2 s_1}{\partial z_1 \partial z_k} \Big/ \frac{\partial^2 s_1}{\partial z_1 \partial x_k} = \frac{\beta}{\alpha}$ for any choice of k. Thus, specifically choosing good k to be the good with the highest value of x works both in our baseline model and when unobservables are instead revealed by search rather than being part of expected utility. We estimate this version of the model in our empirical application of Section 8 and find that the results are robust. In Appendix A.6, we also show that it is possible to test the hypothesis that ϵ is only revealed via search. Indeed, under this hypothesis, we show that the ratio of first derivatives $\frac{\partial s_1}{\partial z_k} \Big/ \frac{\partial s_1}{\partial x_k}$ also recovers $\frac{\beta}{\alpha}$ when k is the good with the highest value of x, so that $\frac{\partial s_1}{\partial z_k} \Big/ \frac{\partial s_1}{\partial x_k} = \frac{\partial^2 s_1}{\partial z_1 \partial z_k} \Big/ \frac{\partial^2 s_1}{\partial z_1 \partial x_k}$. Instead, if ϵ is visible prior to search, in general we have $\frac{\partial s_1}{\partial z_k} \Big/ \frac{\partial s_1}{\partial x_k} \neq \frac{\partial^2 s_1}{\partial z_1 \partial z_k} \Big/ \frac{\partial^2 s_1}{\partial z_1 \partial x_k}$. Thus, testing equality of these two ratios provides a test of the null hypothesis that ϵ is only revealed via search.

3.4 Outside option

So far, we have implicitly assumed that all products are "inside products," in the sense that they each have x and z attributes observed by the researcher. However, in many settings, one may want to model an outside option, which corresponds to choosing none of the products for which attribute data is available. Further, the literature often assumes that consumers know the utility of the outside option for free.[13] First, note that Lemma 1 still holds in this case: if consumers search good 1 (now defined as the *inside* good with the highest value of $\tilde{z}$), they always choose the utility-maximizing good (which could be the outside option). As a consequence, our result can still be applied to estimate β, as we formally show in Appendix A.9.

The presence of the outside option can rationalize the pattern whereby a consumer is faced with a set of options and chooses to not search any of them. In our model, this happens when the utility from the outside option, U_{i0}, is sufficiently high: $g_i(x_j, \epsilon_{ij}, U_{i0}) \leq 0$ for all j. This creates a complication when estimating the distribution of ϵ: consumers might mistakenly choose not to search any of the inside goods, implying that even when we condition on $\tilde{z}_j = \tilde{z}$ for all j, we cannot use full information arguments to trace out the distribution of ϵ. Following many papers in the search literature, we rule this out by assuming that consumers can search one inside good costlessly (e.g., see Hortaçsu and Syverson (2004) and the review in Ursu, Seiler, and Honka (2023)). This assumption implies that at least one of the inside goods is always searched and can be thought of as a way to restrict attention to the population of consumers who are sufficiently interested in the product category.

Finally, in a model with an outside option, one would typically want to allow for a constant term shifting the utility of the inside products relative to the outside option. This can be captured by including among the x variables a dummy for the inside products.

3.5 Expectations parameter γ_1 unknown to the researcher

Our main results in Section 2.2 assumed that the researcher knows the parameters governing consumers' expectations on z. We now consider the case in which the researcher is not willing to assume a value of γ_1. While studying identification of the way consumers form expectations is beyond the scope of this paper, we show that we can still gain some traction under additional assumptions. Specifically, suppose that the sign of γ_1 is known (e.g., higher priced goods are of higher quality). Without loss, we assume $\gamma_1 > 0$. In addition, suppose that there exist choice sets in which a good has both the highest value of z and the lowest value of x. Even when γ_1 is unknown, this good is known to maximize the residual $\tilde{z}_j = z_j - \gamma_1 x_j$ and we label it by 1. With good 1 defined appropriately, the argument from Lemma 2 shows that second derivatives with respect to z_1, z_j, x_j for $j \neq 1$ identify β/α, which just equals β given the normalization $\alpha = 1$. Note that, unlike in the case with γ_1 known, now we cannot in general recover the distribution of ϵ since this requires varying $\mathbf{x}$ while keeping $\tilde{z}_j$ — which is a function of the unknown γ_1 — fixed at a common value for all j. Therefore, this result is not sufficient to simulate choices with full information. However, comparing the ratio of second derivatives to the ratio of first derivatives does allow us to conduct tests for full information as discussed above.

[13] Alternatively, if one assumes that consumers need to search in order to uncover the utility of the outside option, then the latter is no different than any other good and thus our arguments from Section 2.2 immediately apply as long as there are at least two "inside" goods, i.e. goods for which variation in x and z is available in the data.

4 Discussion

4.1 Discussion of Search Model Assumptions

As discussed above, there are several microfoundations for the first assumption. For example, in the Weitzman (1979) search model, consumers search goods in order of reservation utility, which is a function of the product attributes that are visible prior to search, the distribution of the hidden attribute $\tilde{z}_j$, and search costs. If $\tilde{z}_j$ is i.i.d. across goods and consumers have the same search cost for all goods, then consumers will search in order of expected utility (see Example 1). Still, there are at least three reasons this argument might fail: first, there may be more uncertainty about the hidden attribute for some goods than others, and this might lead individuals to search such goods first. Second, unobservables might be correlated across goods, so that, e.g., learning good news about good 1 might cause one to positively update about good 2 and choose to search it before good 3 even if $EU_{i3} > EU_{i2}$. Third, search costs might vary across goods, meaning that consumers prefer to search goods with lower search costs first even if they exhibit lower expected utility than other products.

While the restriction that priors be i.i.d. and search costs be constant across goods is sufficient for Assumption 2(i), it is not necessary. Indeed, our baseline result with $\gamma_1 \neq 0$ allows for non-i.i.d. priors about the unobserved attribute z. Alternatively, priors may be heterogeneous but consumers may be unsophisticated and fail to take into account option value, as in the directed cognition model studied in Gabaix, Laibson, Moloche, and Weinberg (2006). Consumers searching for a laptop online may enter some attributes into a search function and look at the items which rank highly according to those attributes without regard for whether a lower item is worth searching first because its value is more uncertain despite its lower average utility. Such examples also raise the natural concern that in many settings, factors like the order in which items appear in search may impact search costs separately from expected utility. Applications in the marketing literature often allow search costs to vary with observable attributes, such as the position of a good in search (e.g., Ursu (2018)). As discussed in Section 3.2, we extend our main result to allow for these violations of our expected utility assumption by considering cases where some observable attributes impact search but not utility.

Our second assumption on search is that consumers search good j if and only if $g_i(x_j, \epsilon_{ij}, \bar{u}) \geq 0$ where $\bar{u}$ is utility in hand; we also impose the natural restriction that one is (weakly) less likely to search as $\bar{u}$ increases. This assumption is satisfied in most search models we are aware of in the literature, including Weitzman search, satisficing, simultaneously searching all goods with expected utility above a threshold, random search, and directed cognition. One exception is a model in which consumers search simultaneously as in Chade and Smith (2006). This model would violate the assumption because the function g_i that determines whether i searches good j cannot be written only as a function of x_j and ϵ_{ij} since it will depend on the expected utility of all goods. We show in Appendix A.7 that our methods can be extended to accommodate one version of this model based on Honka, Hortaçsu, and Vitorino

(2017). We also investigate the robustness of our approach to a violation of this assumption in the simulations of Appendix D.

Our third assumption, that consumers choose the good which maximizes utility among searched goods, embeds two separate ideas. The first is that consumers do not choose a good they have not searched; this is natural in contexts such as e-commerce, where consumers typically have to open a product's page in order to add it to their carts. The second restriction is that consumers maximize utility given the information available. This could be relaxed by specifying a positive utility function that allows for consumer errors; as long as consumers maximize that positive utility function, the weight that they would attach to the hidden attribute given full information will be revealed. It is then up to the researcher whether to take this weight as the normative benchmark or whether to use some external standard.

The fourth assumption again nests two pieces: (i) that only the value of z_j is unknown prior to search, which we relaxed in Section 3.3, and (ii) that search reveals all information about the hidden attribute. This assumption is natural in settings where z_j is fully observed to the econometrician, as in our case, but is not always plausible. For instance, if the hidden attribute is school value added, a consumer who searches more may learn about test scores and graduation rates, but these are (imperfect) signals of the underlying variable. There is a literature on consumer (Bayesian) learning which models more explicitly the case when search is not fully informative (see Erdem and Keane (1996), Ackerberg (2003), Crawford and Shum (2005), Ursu, Wang, and Chintagunta (2020), among others).

While our assumptions are not without bite, they subsume a range of search protocols and thus are less restrictive than fully specifying a structural search model. Still, one might wonder if they will hold in empirically relevant settings. We address this in two ways. First, in Section 4.3, we show that the assumptions on the search process can be tested based on the same data required for estimation of the model. Second, we apply our approach to data from a lab experiment (Section 7), where we are able to successfully recover preferences based on data with imperfect information, as well as observational data from Expedia (Section 8), where our testing approach correctly identifies the attribute that is not immediately visible to consumers.

4.2 What type of data is required?

The key input to our approach is a flexible estimate of s_1, the choice probability for the good with the highest $\tilde{z}$ as a function of the x and z attributes of all the products in the choice set. As in any discrete choice model, this requires enough variation in the x and z attributes across choice instances.

Standard full-information arguments assume that s_1 can be estimated at a choice set $(\mathbf{x}, \mathbf{z})$ and at a choice set that is identical except that z_j is perturbed by a small amount for some good j. Comparing these two markets nonparametrically pins down the first derivatives of the choice probability functions with respect to z_j. Similarly, nonparametric identification of $\frac{\partial s_1}{\partial x_1}$ requires a third market where x_1

varies holding fixed all other attributes. Thus, recovering $\frac{\partial s_1}{\partial z_1}/\frac{\partial s_1}{\partial x_1}$ requires at least three markets. The required variation in full-information models can come from individual or aggregate data, and can come from cross-sectional variation across markets or panel variation within markets over time. When products have fixed attributes, one typically makes an exchangeability assumption, i.e. that the distribution of the ϵ_{ij} shocks is the same regardless of the identity of product j; variation in product availability across markets then generates the desired variation in $(\mathbf{x}, \mathbf{z})$ even if those attributes are fixed for any given good. The exchangeability assumption likewise accommodates the case where the identity of good 1 (i.e., which good has the largest value of $\tilde{z}$) varies across markets.

In our approach, all of the above still applies – the only difference is that we need more data than under full information since we require second derivatives rather than first derivatives. Specifically, because we take derivatives with respect to three arguments, we need at least six markets to compute our key ratio.[14] Of course, in practice we typically do not observe such clean variation – usually, multiple elements of $(\mathbf{x}, \mathbf{z})$ move simultaneously. Just like in full-information models, this is where parametric restrictions on the choice probability functions are helpful, and data from additional markets or time periods can reduce the dependence on specific parametric assumptions.

One exception to the argument above is the case with endogenous product attributes (Section 3.1). In this setting, just like in full-information models, one needs data from many markets in order to recover the market-level unobservables that are correlated with the endogenous attributes (see Berry and Haile (2014) for the case with aggregate data and Berry and Haile (2020) for the case with individual-level data).

4.3 Testing Search Model Assumptions with and without Observable Search

Our analysis so far has proceeded as if search were not observed; that is, we observe final choices as a function of $\mathbf{x}$ and $\mathbf{z}$ but we do not observe which specific goods were searched. Datasets increasingly contain some information on what is searched: for example, in online clickstream data, one observes not only which product was purchased, but also which products were clicked on en route to purchase (e.g., Ursu (2018)). In many settings, it is plausible to assume that such clicks reveal which products were searched.

Can preferences be identified without resorting to our approach or an explicit search model in these cases? One might speculate that our identification results would be unnecessary; given data on which products were searched, perhaps preferences can be estimated by applying standard methods to the set of searched products without any of the assumptions we require here. However, this is not generally the case because the unobservable component of utility may also drive the search decision. In such

[14] Given a baseline market, recovering $\frac{\partial^2 s_1}{\partial z_1 \partial z_j}$ requires perturbing z_1 alone, z_j alone, and z_1 and z_j jointly. Similarly, recovering $\frac{\partial^2 s_1}{\partial z_1 \partial x_j}$ requires perturbing z_1 alone, x_j alone, and z_1 and x_j jointly. Thus, we need five markets plus the baseline market.

cases, goods with undesirable observables that are searched likely have an especially high realization of ϵ. Thus, it may appear from choice probabilities conditional on search as though the observable attributes are not so bad when in practice, individuals dislike those attributes but this dislike is offset by a large ϵ.[15]

Once our approach is used to identify preferences, clickstream data can be used to conduct additional overidentifying tests. Lemma 2 shows that one can identify the preference parameters as well as the distribution of the ϵ_i shocks. Given this and data on the set of searched goods $\mathcal{G}_i$ — including its size $|\mathcal{G}_i|$ — we can thus compute:

$$P(j \in \mathcal{G}_i|\mathbf{x}, \mathbf{z}) = \sum_k P(|\mathcal{G}_i| = k|\mathbf{x}, \mathbf{z})P(j \in \mathcal{G}_i||\mathcal{G}_i| = k, \mathbf{x}, \mathbf{z}) \tag{11}$$

since the first probability on the right-hand side is observed and the second is pinned down by the model. Specifically, if k goods are searched, the model says that those k goods must be those with the highest expected utility $EU_{ij} = \beta\gamma_0 + (\alpha + \beta\gamma_1)x_j + \epsilon_{ij}$, and this probability can be computed given identification of the preference parameters and of the distribution of ϵ_i. Checking the right-hand side of (11) against the observed search probabilities provides a test of the model.

Even when we do not observe auxiliary information on which goods are searched, the assumptions in our model can be jointly tested by checking whether the observed choice probabilities are consistent with bounds implied by the estimated preferences and assumed search rule. To construct an upper-bound on choice probabilities, note that a good j cannot be chosen if there is an alternative good with higher expected utility and higher utility. Thus, we have:

$$s_j(\mathbf{x}, \mathbf{z}) \leq 1 - P(U_{ik} \geq U_{ij} \text{ and } EU_{ik} \geq EU_{ij} \text{ for some } k) \tag{12}$$

The latter probability can be directly computed from knowledge of preferences and the distribution of ϵ. To construct a lower-bound, note that if good j maximizes both utility and expected utility, then it

[15] A second reason unobservable components of utility might impact search is if preferences are unobservably heterogeneous (random coefficients). Even if search does not depend on ϵ, preferences cannot generally be recovered using only conditional choices unless IIA is satisfied. To see why heterogeneous preferences create a problem, imagine products have quality ratings from 1-5. There are two types of consumers, one type that cares about quality and one type that does not. The type that cares about quality is indifferent about quality over the 4-5 range, but values quality over the 1-4 range sufficiently that quality differences outweigh any other differences observable to consumers. Suppose that quality is observable to consumers (x) but price is only observed conditional on search (z). Quality-conscious consumers only search goods with quality of at least 4. Other consumers will search all goods. If we estimate preferences conditional on search, we will wrongly conclude that no one cares about quality: quality conscious consumers don't care about quality given the goods they have searched (quality ranging from 4-5) and non-quality conscious consumers don't care about quality at all. To estimate preferences correctly, we would have to jointly model the decision of which goods to search and preferences conditional on searching. Thus, with heterogeneous preferences, the existing literature requires specifying a search model in order to estimate preferences even when search is observed in the data. Our approach avoids the need to do this under the assumptions we have outlined. We relegate this point to a footnote because the estimation of high-order derivatives required to identify random coefficients models in our approach may often be impractical.

will be chosen. This implies:

$$s_j(\mathbf{x}, \mathbf{z}) \geq P(U_{ij} \geq U_{ik} \text{ and } EU_{ij} \geq EU_{ik} \text{ for all } k) \tag{13}$$

Once again, the probability on the right can be computed given knowledge of preferences and the distribution of ϵ. Checking whether our estimated choice probabilities are consistent with these bounds provides a test of the model.

Finally, our model is overidentified. For example, in our baseline case with linear utility and homogeneous preferences, $\frac{\partial^2 s_1}{\partial z_1 \partial z_j} / \frac{\partial^2 s_1}{\partial z_1 \partial x_j} = \beta$ for all alternative goods $j \neq 1$ and values of $(\mathbf{x}, \mathbf{z})$ at which the derivative in the denominator is nonzero. This provides a number of overidentifying restrictions which could be used to further test the model.

4.4 Which Counterfactuals Can Our Approach Address?

Here, we discuss the class of counterfactual questions that can be addressed using our method. We start from applications that do not require recovering the distribution of search costs.

Full Information Counterfactuals One important class of counterfactuals asks: how would consumers choose if search costs were reduced? The most natural counterfactuals in our baseline case involve directly informing consumers about the hidden attribute. These counterfactuals are natural in our setting because the hidden attribute is observable to the econometrician.[16] In these cases, knowing preferences is sufficient to simulate how information would impact choices without a structural search model, as we demonstrate in our lab and field experiments. In settings like Hastings and Tejeda-Ashton (2008) or Allcott and Taubinsky (2015) where experimenters fully inform consumers about attributes of goods which were previously accessible at a financial or cognitive cost, our approach can be used to forecast the impact of interventions *before they are conducted*. Additionally, in Appendix G we show how to quantify the welfare gains from more informed choices with an additional separability assumption. Of course, since our approach does not commit to a specific model of search, it does not speak to the gains directly stemming from reduced search costs. In this sense, the estimated increase in welfare can be viewed as a lower bound on the total gains from an information intervention. Estimating the reduction in search costs requires either fully specifying a search model and recovering the cost distribution (we discuss this more at the end of this subsection) or using some auxiliary data on, e.g., time spent searching and value of time.

[16] This can be contrasted with cases where information is only partial and so some search costs likely remain. For example, when unobservables are revealed by search (as in Section 3.3), some information consumers learn upon search is not observable to the econometrician, so informing consumers about the observable component would not eliminate the need to search.

Advertising and Product Design As a second related example, consider a firm trying to understand which features to emphasize in the advertising of a product. Conditional on visible attributes, our results could be used to identify features that consumers value but are not currently always aware of. The firm could use this insight to optimize its advertising strategy, as well as to inform the design of new products (see, e.g., Becker and Murphy (1993) and Bagwell (2007)).

Normative Evaluation of Choices In many counterfactuals where limited information or search costs are not the primary object of interest, one nonetheless is concerned about accurately valuing the attributes of goods. An example is a subsidy for environmentally friendly automobiles. To evaluate such a subsidy, one would conventionally estimate demand and cost parameters in the automobile market (Berry, Levinsohn, and Pakes 1995). If the market were otherwise competitive and efficient, the subsidy might distort choices (creating deadweight loss) but have offsetting externalities. If, however, some consumers are unaware of differences in energy efficiency, the subsidy might redirect them to the cleaner products they would value most if they had more information, meaning that it may be both privately and socially desirable. Our methods can be used to recover whether, prior to imposing the subsidy, consumers are informed about differences in energy efficiency.

We have focused so far on applications where search costs do not need to be recovered. However, our model can also be used to identify search costs given preferences and an underlying structural search model. In Appendix F, we give an explicit example of how search costs can be recovered in a Weitzman model once preferences are known. Intuitively, when preferences are known, we know how consumers would respond to the hidden attribute with zero search costs, and thus we can trace out the distribution of search costs from the observed responsiveness of choice probabilities to the hidden attribute. In Appendix F, we also discuss in more detail questions that can be addressed once search costs are recovered.

5 Estimation

Our identification results show that preferences can be recovered given knowledge of the choice probability function for good 1, denoted by $s_1(\mathbf{x}, \mathbf{z})$. We now discuss how s_1 can be estimated from data on choices and product attributes. We focus on the case with individual-level data (as in the experiment of Section 7 and the application of Section 8). However, our identification approach also applies to aggregate (i.e., market share) data as long as one can consistently estimate the choice probability functions s_j, as discussed in Section 4.2. With individual-level data, our setup implies the following conditional moment restrictions:

$$E\left(y_j - s_j\left(\mathbf{x}, \mathbf{z}\right)|\mathbf{x}, \mathbf{z}\right) = 0 \quad \forall j, \tag{14}$$

where y_j is a dummy variable equal to 1 if a consumer chooses good j. Thus, methods designed to estimate conditional moment restriction models can be used. Of course, the performance of an estimator will depend on how flexibly it captures the derivatives that identify preferences in our approach.

We consider two approaches to estimating $s_1(\mathbf{x}, \mathbf{z})$, and thus its derivatives: (i) an approximation via Bernstein polynomials which is viable when the number of goods and attributes is small; and (ii) a "flexible logit" model which is more ad hoc, but scales better as the number of goods increases.

Both estimation approaches involve choosing tuning parameters; for example, in the Bernstein polynomial approach, one must choose the degree of the polynomial approximation. To deal with this problem, in our experimental analysis, we pre-registered the code used to analyze the experiment. We would encourage other researchers using our method to do the same to avoid concerns about post-hoc tuning to obtain desired results.

5.1 Approximation via Bernstein polynomials

Following Compiani (2022), one can approximate the demand function via Bernstein polynomials. This allows the researcher to impose natural restrictions via linear (and thus easy-to-enforce) constraints on the coefficients to be estimated. Specifically, the class of models considered in this paper satisfies standard monotonicity restrictions in $\mathbf{x}$ and $\mathbf{z}$ (s_j increasing in x_j and z_j and decreasing in $\mathbf{x}_{-j}$ and $\mathbf{z}_{-j}$). In addition, one can consider other constraints, such as exchangeability across goods, which requires demand to only depend on the attributes of the goods, but not their identity.[17] Exchangeability is satisfied if the unobservables entering demand (e.g., preference parameters and shocks, as well as search costs) have the same distribution across goods. We impose both monotonicity and exchangeability in the Bernstein polynomial results reported below. The purpose of these restrictions is twofold. First, they discipline the estimation routine in the sense that they help obtain reasonable estimates of quantities of interest (e.g., negative price elasticities). Second, they help partially alleviate the curse of dimensionality that arises as the number of goods increases. The coefficients in the Bernstein approximation of s_j can be estimated by minimizing a GMM objective function based on the restrictions in (14) subject to the constraints. In Appendix D, we report results from numerous simulations with a variety of data generating processes which suggest that this estimation approach performs well with a small number of goods (2 or 3) when the assumptions of our model are satisfied. It also consistently outperforms standard logit estimates in simulations where our assumptions are violated. This approach has the advantage of providing a nonparametric approximation to the choice probability function s_1, but it suffers from a curse of dimensionality that makes it very data-demanding for ≥ 4 goods.

[17] See Compiani (2022) for a formal definition of exchangeability.

5.2 "Flexible Logit"

As the number of goods increases, nonparametric methods face a curse of dimensionality, and thus it becomes necessary to place some additional structure on the problem. In this section, we develop one such approximation which performs well in simulations for a larger number of goods.[18]

As discussed in more detail in Appendix E, conventional full-information models typically impose strong restrictions on the structure of the derivatives of choice probabilities. For example, in a multinomial logit model, it is always the case that $\frac{\partial^2 s_1}{\partial z_1 \partial z_j} / \frac{\partial^2 s_1}{\partial z_1 \partial x_j} = \frac{\partial s_1}{\partial z_j} / \frac{\partial s_1}{\partial x_j}$. Since our test for full information is based on detecting discrepancies between these two ratios, a logit model will clearly not be suitable. To allow for the required additional flexibility, we let the mean utility for good 1 depend directly on attributes of rival goods as follows:

$$ax_1 + b_1 z_1 + \sum_{k \neq 1} (\eta_k w_{z1k} z_k + \eta_{2k} w_{x1k} x_k + \rho_k w_{z2k} z_k z_1 + \rho_{2k} w_{x2k} x_k z_1) \tag{15}$$

where w_{z1k}, w_{x1k}, w_{z2k} and w_{x2k} are known weights, and a, b_1, η_k, η_{2k}, ρ_k and ρ_{2k} are coefficients to be estimated. Further, we let the mean utility for $k \neq 1$ take the standard form $ax_k + bz_k$. If we set the weights equal to 1, this would be a conventional logit model where the utility for good 1 is allowed to depend flexibly on the attributes of good 1, as well as rival attributes interacted with the attributes of good 1. This is the idea behind the "universal logit" model (McFadden 1981; McFadden 1984) and thus our method can be viewed as an extension of this designed to target our second derivatives of interest. Specifically, approximating choice probabilities with estimates from a standard logit model, it is possible to choose the weights as a function of choice probabilities so that the structural parameter of interest (β/α) is a closed form function of the estimated coefficients (we show this formally in Appendix E). More precisely, with weights chosen appropriately:

$$\frac{\partial^2 s_1}{\partial z_1 \partial z_j} \Big/ \frac{\partial^2 s_1}{\partial z_1 \partial x_j} = \frac{-b + \eta_j + \rho_j}{-a + \eta_{2j} + \rho_{2j}}, \tag{16}$$

where the coefficients on the RHS are recovered directly by estimating the model based on equation (15). The weights are estimated as functions of choice probabilities and coefficients from a naive logit model.[19]

We note that the parameters in (15) do not have a causal interpretation (i.e., we are not positing that the actual utility of good 1 depends on the attributes of good k for $k \neq 1$). Instead, (15) is simply

[18] For example, in our application in Section 8, we have 10 products per choice set and flexible logit involves 46 parameters in total. In contrast, a fully nonparametric approach would involve a much larger number of parameters (e.g., even if we modeled s_1 as a polynomial of up to degree 1 in each of the arguments $(\mathbf{x}, \mathbf{z})$, we would have $2^{20} = 1,048,576$ parameters).

[19] Specifically, we assume $u_{ij} = \alpha^* x_j + \beta^* z_j + \epsilon_{ij}$ and compute the implied choice probabilities s_j^*. Then: $w_{z1j} = w_{x1j} = \frac{s_j^*}{1-s_1^*}$ and $w_{x2j} = w_{z2j} = \beta^* \frac{(1-2s_1^*)s_j^*}{1-s_1^*} (1 + \beta^*(1-2s_1^*)z_1)^{-1}$.

a flexible function of $(\mathbf{x}, \mathbf{z})$ that we found captures the second derivatives of s_1 well. In Appendix D, we show that the flexible logit performs extremely well in simulations for a variety of data generating processes. For three DGPs satisfying the assumptions of our model, conventional logit estimates are biased, but flexible logit confidence intervals include the true values. For a fourth DGP violating the assumptions of our model, flexible logit has a small bias with a large number of goods, but is consistently less biased than the standard logit estimates. This being said, the approach is somewhat ad-hoc and we welcome more formally validated approaches for optimizing the trade-off between flexibility and scalability in estimating our second derivatives of interest.

In subsequent applications, we have found that the performance of flexible logit deteriorates when the impact of z on choices is small in the data, leading confidence intervals to explode. This is intuitive regardless of how our model is estimated: if $\frac{\partial s_1}{\partial z_1}$ is small, then $\frac{\partial^2 s_1}{\partial z_1 \partial z_j} / \frac{\partial^2 s_1}{\partial z_1 \partial x_j}$ is difficult to estimate with any precision unless the dataset is extremely large. Our approach is thus better suited to settings where a potentially hidden attribute has a meaningful impact on choices in the raw data (e.g., school choice as a function of high school test scores).

6 Practitioner's Guide

Table 1 provides guidance for applied researchers seeking to use our approach. As in the previous section, we focus on the case with individual-level data, although these procedures could be adapted to the case with aggregate market-level data. We also assume the researcher is willing to fix the variance of ϵ via parametric assumptions on its distribution (e.g., assume it is Gumbel distributed), so that α cannot be normalized without loss and instead needs to be estimated. Further, we focus on the case with continuous attributes and discuss how to handle discrete attributes in Appendix A.8. Appendix A.11 outlines how to adapt the approach to the case where the researcher wishes to let x enter utility nonlinearly.

The estimation routines outlined in Table 1 are likely to lead to slower-than-parametric convergence rates. This is because our parameters of interest are estimated by evaluating nonparametric functions at specific points and/or using a slice of the data (e.g., for the estimation of α in step 2 (c)).[20] We are however reassured by the fact that the methods perform well in both our experiment and our empirical application.

[20] This step is akin to estimation for varying coefficients models (Cai, Fan, and Li 2000).

Table 1: Steps for Estimation

1. For each choice set in the data, identify good 1: the good with the highest $\tilde{z}$ (or, if z is vector-valued, the good with the highest weighted $\tilde{z}$-index obtained via the procedure in Appendix A.3).
2. For the baseline model of Section 2.2, follow these steps:
   (a) If the number of goods is 2 or 3, estimate the model nonparametrically:
      i. Use equation (14) to estimate the choice probability function s_1 using Bernstein polynomials subject to monotonicity and exchangeability restrictions (Compiani 2022).
      ii. Compute estimates of $\frac{\partial^2 s_1(\mathbf{x},\mathbf{z})}{\partial z_1 \partial z_j}$ and $\frac{\partial^2 s_1(\mathbf{x},\mathbf{z})}{\partial z_1 \partial x_j}$ for all $j \neq 1$ and all choice sets $(\mathbf{x}, \mathbf{z})$ in the data.
      iii. Take the ratio of a trimmed mean of $\frac{\partial^2 s_1(\mathbf{x},\mathbf{z})}{\partial z_1 \partial z_j}$ to a trimmed mean of $\frac{\partial^2 s_1(\mathbf{x},\mathbf{z})}{\partial z_1 \partial x_j}$ (both means are over choice sets $(\mathbf{x}, \mathbf{z})$). Further average these ratios over $j \neq 1$ to obtain an estimate of $\frac{\beta}{\alpha}$.
   (b) If the number of goods is greater than 3, estimate the model using flexible logit (Section 5.2):
      i. Create initial naive estimates (α^*, β^*) and associated choice probabilities s_j^* by estimating a naive logit model. Use these to construct the weights w_{ij} as described in Appendix E.
      ii. Estimate equation (15) given these weights to recover the coefficients: $a, b, (\eta_k, \eta_{2k}, \rho_k, \rho_{2k})_{k \neq 1}$.
      iii. Use equation (16) to compute $\frac{\partial^2 s_1}{\partial z_1 \partial z_j} / \frac{\partial^2 s_1}{\partial z_1 \partial x_j}$ and take an average across j (or impose coefficient restrictions such that $\eta_j, \eta_{2j}, \rho_j, \rho_{2j}$ are the same for all j). This gives an estimate of $\frac{\beta}{\alpha}$.
   (c) Recover β as follows:
      i. Estimate α via standard full-information models (e.g., logit), only using choice sets where the variance of $\tilde{z}_j$ across j is below a cut-off.
      ii. Multiply the estimate of β/α from step 2(a) or 2(b) by the estimate of α from step 2(c)i to obtain an estimate of β.
3. If some attributes are endogenous, modify step 2(a)i to incorporate instruments (as in Compiani (2022)). Then follow steps 2(a)ii-iii where now the choice probability s_1 is a function not only of $(\mathbf{x}, \mathbf{z})$, but also of the unobservables causing endogeneity ξ (which can be fixed once the structural demand function is estimated).
4. If the data contains variables that are likely to affect search but not utility (e.g., rankings r on a webpage), drop all choice sets where consumers choose a product ranked lower than product 1 ($r_j < r_1$). Then, follow step 2 using the set $\mathcal{R} = \{j : r_j \geq r_1\}$ as the choice set and including r among the x variables. In step 2(c)i, also condition on choice sets where the variance of r_j across j is sufficiently low.
5. To allow for the possibility that ϵ is only revealed via search, choose product j in step 2(a)ii or 2(b)iii to be the good with the best value of x (which could coincide with good 1).
6. Optional:
   (a) In order to increase power, repeat the previous steps for various choices of the x variable (if available) and take a weighted average of the resulting β estimates with weights proportional to their precision.
   (b) Test the model by checking whether the estimated choice probabilities lie within the bounds derived in Section 4.3.

Note: We recommend using simulations to choose the following tuning parameters prior to estimation: the polynomial degree in step 2(a)i, the amount of trimming in step 2(b)iii, and the variance cutoffs in steps 2(c)i and 4.

7 Experimental Validation

Our identification proof and simulation results show that preferences can be estimated regardless of whether consumers are fully informed, provided consumers search in a way that is consistent with our assumptions. Of course, this does not tell us whether those assumptions are likely to be satisfied in practice.

In this section, we test in a lab experiment whether we can recover preferences in a setting where consumers engage in costly search. Unlike in our simulations, the search protocol is unknown to us and not restricted to satisfy all of our model assumptions. In particular, while we enforce the assumption that consumers must search a product in order to choose it, we do not constrain the order in which the various options are searched nor the rule that determines when to stop searching. We nonetheless show that we are able to correctly recover preferences using our "search-robust" estimation technique.

7.1 Set-up

We selected 1,000 books for sale on Amazon Kindle chosen from a wide variety of genres. For each book, we observe its average rating on the site "Goodreads.com" as well as the average rating from Amazon.com, the number of reviews on Goodreads, and the price of the book for Amazon Kindle.

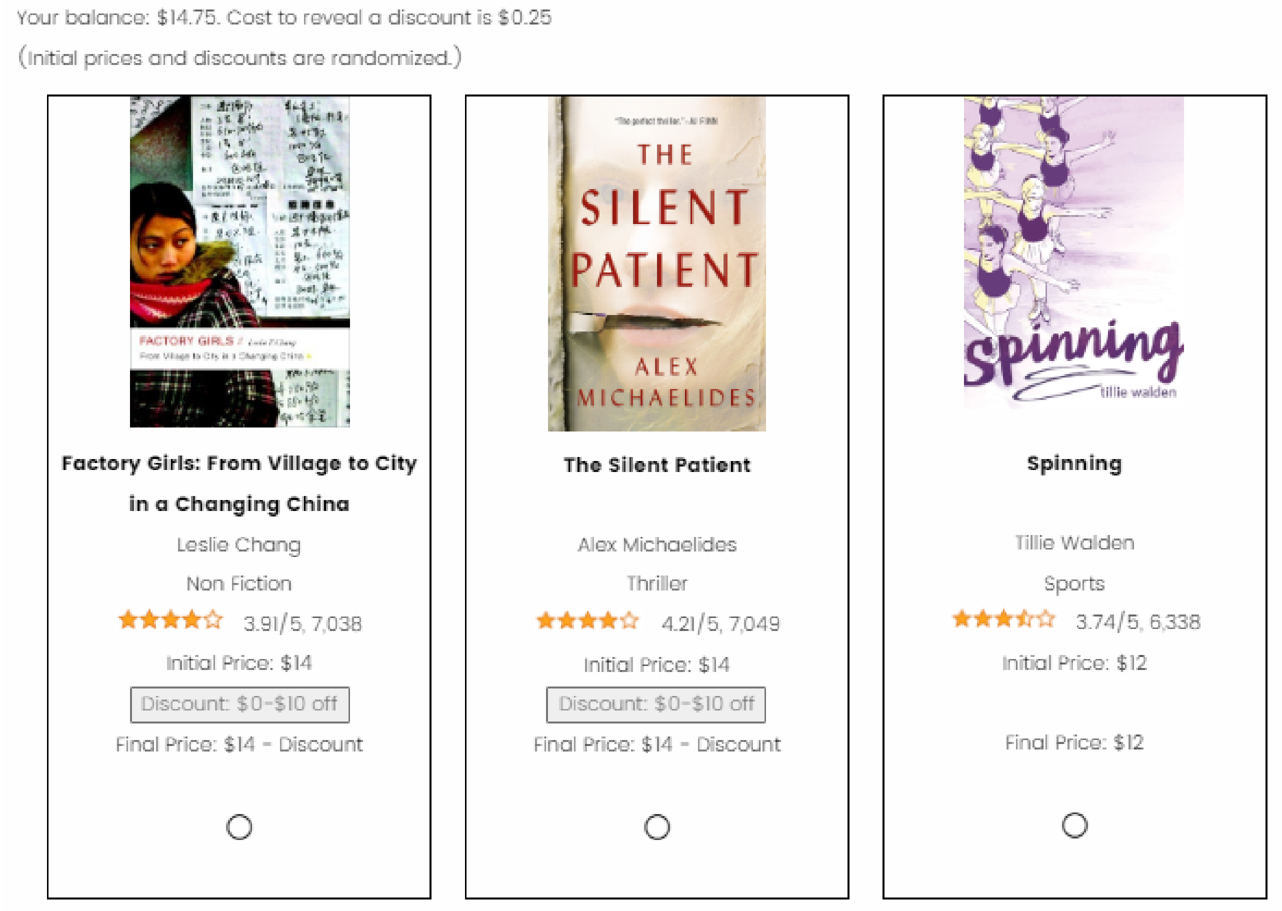
In our experiment, conducted via Connect CloudResearch,[21] each participant made 20 choices from sets of 3 randomly selected books. In each choice, participants had to select a book, i.e. they had no outside option. For all books, participants could see a photo of the cover, the title, author and genre, as well as the Goodreads rating and the number of ratings. Prices were randomized to integers from \$11-\$15 (equally likely). All books were then further discounted by an integer amount uniformly drawn from \$0-\$10 (and participants were informed of this). All users were given a \$15 bank at the start of each choice, from which any costs incurred were deducted. There were a total of 196 participants, yielding 3,920 choices.

The discount is our key variable of interest. For 5 of the 20 choices, users could see all discounts and thus could see the net price of all options at no cost. For 15 of the 20 choices, discounts were hidden and users had to pay a cost to see the discount for any given book.[22] The cost per click was constant for each user across the 15 choices, and randomly chosen from $\{\$0.10, \$0.25, \$0.35, \$0.50\}$. For the 15 choices with hidden information, users could only choose books after they clicked to reveal the discount and had to choose at least one book. One of the 20 choices made by each user was randomly chosen to be realized, and users received the chosen book as well as any money left over from the original \$15. Figure 1 shows a sample product selection screen from a choice where discounts were hidden. In this

[21] We preregistered the experiment and estimation codes at the platform OpenScienceFramework: https://doi.org/10.17605/OSF.IO/D4XA2.

[22] The full information and costly information choice situations were randomly ordered, so that the 5 "full information" choices were intermixed with the costly information choices.

Figure 1: Lab Experiment: Sample Product Selection Screen



case, the user clicked to reveal the discount of the second book and could either choose that book or continue by revealing the discounts for additional books. Note that the user could search books in any order she wished. The average total number of clicks per participant per choice is 1.5.

The 5 choices where all information is revealed are our benchmark for the "truth." The goal is then to test whether the weight on discounts relative to prices that we estimate in the cases where discounts are costly to observe matches the relative weight we see when discounts are visible to everyone (i.e. can we predict informed choices using data from uninformed ones). Further, because both discounts and prices are in dollar terms, and because they are randomized (and thus not signals of quality), there is a second benchmark: if consumers are rational, the weight on discounts and prices should be equal.

We will model choices using the following utility specification:

$$U_{ij} = price_j \cdot \alpha_1 + rating_j \cdot \alpha_2 - discount_j \cdot \beta + \epsilon_{ij} \tag{17}$$

where ϵ_{ij} is i.i.d. type-I extreme value and accounts for any aspect of consumers' tastes for books (based on the title, image, author or genre) not summarized by the price, discount and rating variables. Fully informed and rational consumers should have $\alpha_1 = \beta$. Our goal will be to show that we can recover these fully informed preferences using the choices of beneficiaries for whom revealing discounts is costly.[23]

7.2 Estimation Results

Columns 1 and 2 of Table 2 show results from estimating a standard logit model on consumer choices for the 5 choice situations per consumer where all information is revealed (Full Info) and the 15 choice situations where consumers must pay to reveal information (Costly Info), respectively. With full information, consumers place equal weight on prices and (negative) discounts, so they pass our test of rationality. In other words, they care only about the final price of the product. By contrast, when discounts are costly to reveal, the coefficient on the discount variable in the standard logit model is attenuated (the "Costly Info" column). This is because consumers are insensitive to variation in discounts for books they do not search. The ratio of the two coefficients is 1.094 in the full information treatment and 0.665 in the costly information treatment, consistent with the attenuation bias established in Lemma 3.

Following Section 5.1, we estimate the demand function $s_1(\mathbf{x}, \mathbf{z})$ via Bernstein polynomials. The exact procedure is described in Appendix H. Our estimate of β/α_1 is 1.183, and the confidence interval contains the corresponding estimate of 1.094 from column 1 (and the theoretical estimate of 1), and is sufficiently tight to exclude the standard logit estimates in the costly information treatment. Besides estimating β/α_1, we need to directly recover the α coefficients. Consistent with the proof of Lemma 2, we compute these by estimating a logit model using only choice sets where the variance of the discount across goods is in the bottom quintile. The results are reported in column 3 of Table 2, along with the value of β implied by our estimates of α_1 and β/α_1. Again, the confidence intervals include the full information values. In other words, using data only on choices when information is costly, we successfully recover informed preferences.[24]

Having recovered all preference parameters, we can compute how information will change behavior and choice quality. Using only data on choices when search is costly, our model predicts that, on average, full information consumers would save \$0.71 per choice from choosing books with lower discounts. The corresponding number in the data is \$0.64 per choice situation, since consumers in the costly information treatment average discounts of \$5.82, while consumers in the full information treatment

[23]Note that, given our parametric assumption on the distribution of ϵ, we are not free to normalize any of the α coefficients here.

[24]To show validity of the nonparametric bootstrap for the case where the constraints are not binding asymptotically, one can combine the general framework of conditional moment models with different conditioning variables (Ai and Chen 2007) with the results in Section 5 of Chen and Pouzo (2015). An interesting avenue for future research is to explore the case where the constraints are binding asymptotically.

Table 2: Standard Logit and Second Derivative Estimation Results

	Standard Approach		**Our Approach**
Variable	**Full Info**	**Costly Info**	**Costly Info**
Price	-0.224***	-0.182***	-0.230***
	(0.031)	(0.017)	(0.032)
Discount (-)	-0.245***	-0.121***	-0.272***
	(0.015)	(0.008)	(0.055)
Rating	0.350**	0.471***	0.588***
	(0.166)	(0.091)	(0.171)
Discount (-) / Price	1.094***	0.665***	1.183***
	(0.156)	(0.075)	(0.214)
N	980	2940	2940

Note: The table shows estimation results from a standard logit model estimated on the full information and costly information treatments in columns 1 and 2, and Bernstein polynomials estimation of the second-derivative ratio on the costly information treatment in column 3. The minus sign indicates that discount is multiplied by -1 so that the coefficient on discount should equal that of price under full information. Standard errors on the discount coefficient and the ratio of the discount and price coefficients are computed using 250 bootstrap draws. *** denotes significance at the 1% level, ** at 5% level, and * at 10%.

average discounts of \$6.46.[25] In other words, we can accurately predict how consumers will respond to information provision *before* the information is provided. We can also compute the dollar equivalent welfare benefits of providing consumers with information. To do so, we take our estimates from column 1 as the normative preferences (i.e., as the correct metric to compute consumer welfare) and calculate by how much welfare changes when consumers go from making partially uninformed choices to fully informed choices using the approach in Appendix G. We then repeat this exercise using the estimates from column 3 as the normative preferences. We estimate an average welfare gain of \$0.18 per choice based on column 1 and of \$0.25 based on column 3. Thus, our model again yields results that are fairly close to those coming from the "true" fully informed choices in the data.[26]

As in most real-world settings, expected utility is not observable to the econometrician in our experiment: while we can see attributes of the goods in question, we do not know how individuals will weigh these attributes, nor do we know their preferences for specific genres or book titles and images (captured by ϵ_{ij} in our model). The assumption that consumers search in descending order of expected utility is substantive and could be violated in numerous ways: users might always reveal discounts for the lowest priced book first or they might search in the order in which books are displayed. Nonetheless, our "robust" estimation approach succeeds in recovering the preferences that consumers reveal in the full information condition. In Appendix I, we report results from the test discussed in Section 4.3, showing additionally that the estimated choice probabilities lie within the upper and lower bounds implied by our expected utility assumption. Thus, we fail to reject the assumption.

8 Field Validation

Our lab experiment demonstrates one setting where our approach correctly identifies hidden attributes and forecasts how consumers will respond to information about hidden attributes. Of course, this leaves open the question of whether we can identify preferences in real-world settings with a larger number of goods, where search costs are implicit and potentially heterogeneous, and where we (as experimenters) cannot strictly control the information available to consumers.

In this section, we investigate these issues using publicly available data from a leading online travel agency, Expedia (Ursu 2018).[27] The data we use includes transactions from 54,648 consumers over an eight-month period between November 1, 2012, and June 30, 2013.[28] At the time, consumers would search for a hotel in a given city, and Expedia would present a list of available options. On the list,

[25]We look at differences in discounts as opposed to final prices paid since the latter are essentially the same in the full information and in the costly information conditions.

[26]Note that the benefits are smaller than the increase in discounts because information induces consumers to be more responsive to discounts, sacrificing some value on unobservable factors. Further, we put the word "true" in quotes because the full information logit may be misspecified even in column 1 due to, e.g., correlation in ϵ across products.

[27]The data is available for download at `https://www.kaggle.com/c/expedia-personalized-sort/data`.

[28]This is the final dataset. Appendix J.1 contains details of data cleaning and variable descriptions.

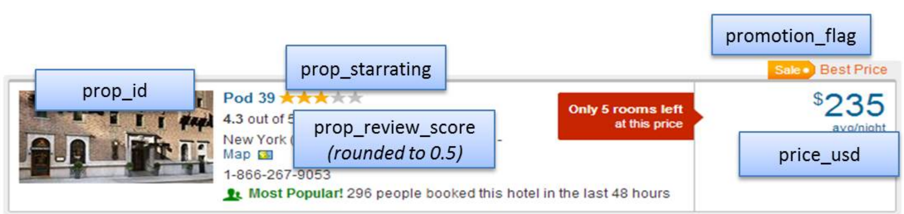
consumers observe a range of hotel characteristics including the price per night, whether the hotel is on promotion or part of a chain, star rating, and the review score.

One attribute, location, is not visible to consumers in the search results but is only visible with additional effort, either clicking on the hotel or clicking to open a map showing unlabeled pins, and then finding the pin of a hotel. The dataset contains a measure of location desirability, but this measure is not visible to consumers in the search results. We thus ask: can our testing approach in Lemma 4 correctly recover that location is a hidden attribute, whereas other attributes are directly visible on the search results page?

Note that while the expected utility assumption is not ex ante unreasonable in this setting, there are several ways that it might be violated. The most obvious is that the ranking of hotels in search results might matter directly, and we allow for this in an extension below. Consumers may ignore the search results and click on a map, where they can see location but not the other attributes of hotels (a very different search process). Or consumers might make errors, such as searching lower priced hotels first with no consideration for how expected location varies with price. While we cannot observe these behaviors directly, we will construct bounds on choice probabilities implied if the expected utility assumption is satisfied.

Table 3 provides summary statistics. Hotels on average charge \$162 per night, with 3.5 stars and a review score of 4 out of 5. 64% of the hotels belong to a chain, and 34% of the hotels display a promotion. Location attractiveness is a score ranging from 0 to 7 designed by Expedia to measure how centrally a hotel is located, what amenities surround it, and other aspects of location desirability. The average hotel has a location score of 3.26. Figure 2 illustrates how hotel characteristics appear to consumers in search. Note that information on features like price, stars, review score, and promotion flag are saliently displayed. However, consumers do not observe the detailed map unless they click, and the quantitative location score is not shown. In order to evaluate the attractiveness of a location, consumers need to spend time and effort to examine the map.

Figure 2: Hotel Characteristics Shown in the Search Impression



We consider choices with the following utility specification:

$$U_{ij} = \tilde{\mathbf{x}}_j \cdot \tilde{\alpha} + \mathbf{x}_j \cdot \alpha + z_j \cdot \beta + \epsilon_{ij} \tag{18}$$

where ϵ_{ij} is type-I extreme value and $\tilde{\mathbf{x}}_j$ is the vector of attributes which we always model as visible,

Table 3: Summary Statistics

	Observations	Mean	Median	SD	Min	Max
Price ($)	546,480	162.11	139.57	92.94	10	1000
Stars	546,480	3.42	3.00	0.91	0	5
Review Score	546,480	4.01	4.00	0.71	0	5
Chain	546,480	0.64	1.00	0.48	0	1
Location Score	546,480	3.26	3.22	1.45	0	7
Promotion	546,480	0.34	0.00	0.47	0	1

including the chain, promotion, and position dummies.

Table 4: Model Specifications

Model	$\mathbf{x_j}$	z_j
I	Price, Stars, Review Score	Location Score
II	Stars, Review Score, Location Score	Price
III	Price, Stars, Location Score	Review Score
IV	Price, Review Score, Location Score	Stars

For the four remaining variables — stars, review score, location score, and price — we estimate four models where one of these variables plays the role of z_j in our model (i.e., it may only be revealed after search), and the other three serve as $\mathbf{x}_j$ (i.e., they are assumed to be known pre-search). Table 4 shows the model specifications. We start by assuming that consumers don't form expectations on the hidden attribute based on the visible attributes, and relax this assumption in a robustness check. We estimate each model using the "flexible logit" approach described in Section 5.2. Since we have several different x variables for each model, in order to increase power, we estimate β by taking derivatives with respect to each individual x variable and then average the results (step 6 of Table 1).[29]

Figure 3 shows the estimates from standard logit and flexible logit for each candidate z variable.[30] Location score is the only variable where we see clear evidence that standard logit is attenuated relative to flexible logit, which is consistent with the fact that location is not immediately available to consumers in the results page. Thus, standard logit tends to underestimate how much consumers value location. On the contrary, for the visible attributes – price, review score and star rating – we find that the flexible logit confidence intervals include the standard logit estimates, and the differences between flexible logit and standard logit estimates are not significantly different from zero.

Given the evidence that location is a hidden attribute, we use our estimates of consumer preferences to compute how information about location will change behavior and choice quality. Table 5

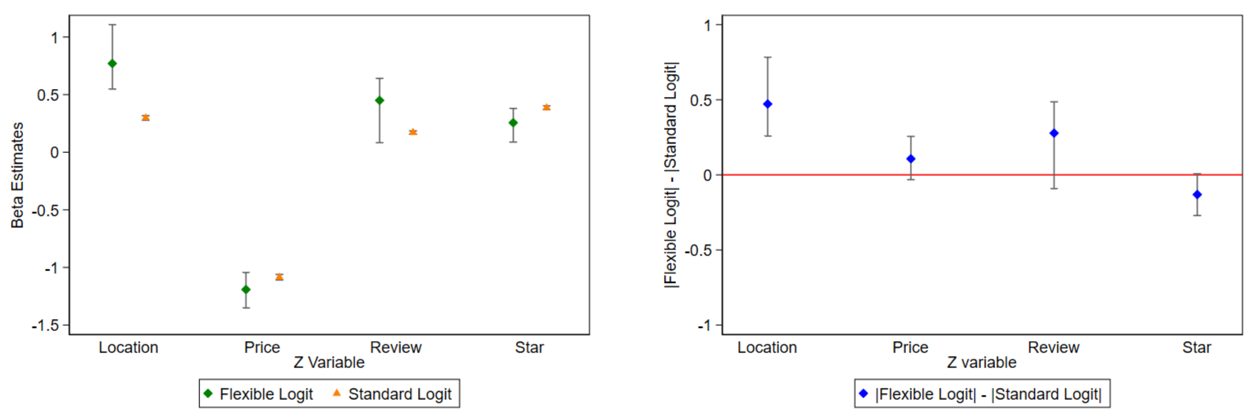
[29] Appendix J.2 contains a step-by-step guide for computing the point estimates and confidence intervals.

[30] Values are reported in Appendix J.3.

Figure 3: Estimation Results

(A) β Estimates

(B) Difference in Magnitude of β Estimates



Note: In Panel A, we report 95% confidence intervals for the coefficient β for different choices of the z variable. In each case, we normalize the coefficient by multiplying it by the standard deviation of the variable. In Panel B, we report 95% confidence intervals for the difference between the absolute value of the normalized β estimate from flexible logit and the absolute value of the corresponding standard logit estimate.

shows the counterfactual results when we make the location score information visible to all consumers. The average location score among transacted hotels increases from 3.32 in the data to 3.41 in the counterfactual scenario where location is fully visible. Further, using the approach in Appendix G, we compute the welfare benefits of providing consumers with location score information. We estimate an average welfare gain of \$2.20 per choice, which is 1.6% of the average transaction price. Interestingly, the average price paid is higher in the status quo relative to the full-information counterfactual, suggesting that the platform is benefiting from consumers' imperfect information.

Table 5: Counterfactual Results

	Status quo	Counterfactual
Average Value for the Transacted Item		
Price (\$)	143.52	137.06
Stars	3.47	3.42
Review Score	4.03	4.03
Chain	0.64	0.64
Promotion	0.40	0.39
Position	4.55	4.91
Location Score	3.32	3.41
Welfare Difference per Choice (\$)	-	2.20

Note: Average value of different attributes for the transacted item in the data (first column) and in the counterfactual scenario where consumers have full information on location (second column). The last row reports the average welfare change from the status quo to the counterfactual.

In Appendix J.4, we conduct a series of additional analyses to check the robustness of the results.

Specifically, we show that our results qualitatively do not change if: (i) we allow consumers to form expectations about location based on the other attributes (i.e., $\gamma_1 \neq 0$); (ii) we let the position of the hotel on the results page affect the order of search as in Section 3.2; (iii) we let the idiosyncratic term ϵ_{ij} be revealed to consumers only after search (Section 3.3). Additionally, (iv) we report results from the test discussed in Section 4.3 and show that the estimated choice probabilities almost always lie within the upper and lower bounds implied by the assumption that consumers search in descending order of expected utility. Further: (v) we check whether consumers who searched good 1 are systematically different from others by comparing the total number of clicks conditional on searching good 1, and find that they do not behave differently from other consumers; (vi) we explore robustness to how much data we use to estimate the coefficients on the visible attributes (which theoretically requires $\tilde{z}_j = \bar{z}$): we find that the estimates from flexible logit are stable as we vary the amount of data, and significantly different from the standard logit estimates; and (vii) we present a test of whether the estimates of the β coefficient on the location score vary across choices of which x variable is used to construct our ratio of second derivatives. We fail to reject the null hypothesis that the estimates obtained from different choices of x variables are equal.

9 Conclusion

We give sufficient conditions to estimate preferences using only data on attributes and choices in cross-sectional or panel data even when consumers must search to acquire information about product attributes. The approach is robust in the sense that it works regardless of whether consumers have full information or engage in search based on a broad class of search protocols. Further, our results can be used to test whether consumers are fully or only partially informed about a given attribute.

Because our conditions allow preferences to be recovered when consumers are imperfectly informed, our results allow a wide range of inquiries that are impossible using conventional methods. First, prior to conducting an information intervention, choice data can be used to estimate counterfactually how consumers would choose were they fully informed. Further, our approach can be used to conduct reliable welfare analyses that are based on consumers' true preferences as opposed to the preference estimates obtained from standard methods assuming full information, which may be confounded by consumers' lack of information.

In many settings, our assumptions may not be exactly satisfied, but these must be assessed relative to the alternatives. The vast majority of empirical work currently makes the often dubious assumption that consumers are fully informed about all attributes of products.[31] Even if one lacks contextual information to support our assumption that consumers search in descending order of expected utility,

[31] We count 350 articles published in the AER, QJE, JPE, ECTA or ReStud since 2015 that estimate discrete choice models. Of these 350, 315 (90%) assume that consumers are fully informed. The list of papers and their classification is available upon request from the authors.

our approach is much weaker than the standard assumption of full information and may be preferable in settings where estimates of preferences are needed to conduct welfare analysis.

Relative to methods based on the full information assumption, the main downside of our approach is that it is more demanding of the data and places limits on the extent of consumer heterogeneity allowed. For example, our identification arguments require that consumers agree about whether the possibly hidden attribute is good or bad and; further, recovering the distribution of preference heterogeneity requires parametric restrictions or very large datasets.

A second shortcoming of our approach currently is the ad hoc nature of the "flexible logit" estimation procedure that we use in large choice sets. We find that this functional form works well in a range of simulations, but we have not stated formal assumptions under which this estimation procedure will recover the true choice probability function. One implication of our theory is that the strong assumptions made in existing parametric models about cross-derivatives rule out realistic behavior. Work clarifying the formal assumptions necessary to relax these restrictions while maintaining scalability in large choice sets would be welcome.

Our assumptions are sufficient for identification but not necessary, raising many questions: are there other conditions aside from the expected utility assumption which permit analogous data-driven identification of consumers who maximize utility? Are there necessary and sufficient conditions for preferences to be recoverable from choice data when consumers have partial information? Finally, our approach could be combined with consideration sets methods to allow for imperfect information at both the attribute and the product level. This may be desirable to assess information interventions that inform both about the attributes of products and about which alternatives exist.

References

Abaluck, J. and A. Adams (2017). What do consumers consider before they choose? identification from asymmetric demand responses. Technical report, National Bureau of Economic Research.

Abaluck, J. and J. Gruber (2009). Choice inconsistencies among the elderly: Evidence from plan choice in the medicare part d program (working paper# 14759). *Cambridge, MA: National Bureau of Economic Research. Retrieved from http://www. nber. org/papers/w14759. doi 10*, w14759.

Abaluck, J. and J. Gruber (2011). Choice inconsistencies among the elderly: evidence from plan choice in the medicare part d program. *The American economic review 101*(4), 1180–1210.

Ackerberg, D. (2003). Advertising, Learning, and Consumer Choice in Experience Good Markets: An Empirical Examination. *International Economic Review 44*.

Agarwal, N. and P. J. Somaini (2022). Demand analysis under latent choice constraints. Technical report, National Bureau of Economic Research.

Ai, C. and X. Chen (2007). Estimation of possibly misspecified semiparametric conditional moment restriction models with different conditioning variables. *Journal of Econometrics 141*(1), 5–43.

Allcott, H., B. B. Lockwood, and D. Taubinsky (2019). Regressive sin taxes, with an application to the optimal soda tax. *The Quarterly Journal of Economics 134*(3), 1557–1626.

Allcott, H. and D. Taubinsky (2015). Evaluating behaviorally motivated policy: Experimental evidence from the lightbulb market. *American Economic Review 105*(8), 2501–38.

Allen, R. and J. Rehbeck (2020). Identification of random coefficient latent utility models. *arXiv preprint arXiv:2003.00276*.

Anderson, S. P., F. Ciliberto, and J. Liaukonyte (2013). Information content of advertising: Empirical evidence from the otc analgesic industry. *International Journal of Industrial Organization 31*(5), 355–367.

Armstrong, M. (2017). Ordered consumer search. *Journal of the European Economic Association 15*(5), 989–1024.

Bagwell, K. (2007). Chapter 28 The Economic Analysis of Advertising. *Handbook of Industrial Organization 3*, 1701–1844.

Barseghyan, L., M. Coughlin, F. Molinari, and J. C. Teitelbaum (2021). Heterogeneous choice sets and preferences. *Econometrica 89*(5), 2015–2048.

Becker, G. S. and K. M. Murphy (1993). A Simple Theory of Advertising as a Good or Bad. *The Quarterly Journal of Economics 108*(4), 941–964.

Berry, S., A. Gandhi, and P. Haile (2013). Connected substitutes and invertibility of demand. *Econometrica 81*, 2087–2111.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, 841–890.

Berry, S. T. and P. A. Haile (2014). Identification in differentiated products markets using market level data. *Econometrica 82*(5), 1749–1797.

Berry, S. T. and P. A. Haile (2020). Nonparametric identification of differentiated products demand using micro data. Technical report, National Bureau of Economic Research.

Branco, F., M. Sun, and J. M. Villas-Boas (2012). Optimal search for product information. *Management Science 58*(11), 2037–2056.

Bronnenberg, B. J., J.-P. Dubé, M. Gentzkow, and J. M. Shapiro (2015). Do pharmacists buy bayer? informed shoppers and the brand premium. *The Quarterly Journal of Economics 130*(4), 1669–1726.

Bronnenberg, B. J., J. B. Kim, and C. F. Mela (2016). Zooming in on choice: How do consumers search for cameras online? *Marketing science 35*(5), 693–712.

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

Cai, Z., J. Fan, and R. Li (2000). Efficient estimation and inferences for varying-coefficient models. *Journal of the American Statistical Association 95*(451), 888–902.

Cattaneo, M. D., X. Ma, Y. Masatlioglu, and E. Suleymanov (2020). A random attention model. *Journal of Political Economy 128*(7), 2796–2836.

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

Chen, X., T. Christensen, and S. Kankanala (2024). Adaptive estimation and uniform confidence bands for nonparametric structural functions and elasticities. *Review of Economic Studies*, rdae025.

Chen, X. and D. Pouzo (2015). Sieve wald and qlr inferences on semi/nonparametric conditional moment models. *Econometrica 83*(3), 1013–1079.

Choi, M., A. Y. Dai, and K. Kim (2018). Consumer search and price competition. *Econometrica 86*(4), 1257–1281.

Compiani, G. (2022). Market counterfactuals and the specification of multiproduct demand: A nonparametric approach. *Quantitative Economics 13*(2), 545–591.

Compiani, G., G. Lewis, S. Peng, and P. Wang (2024). Online search and optimal product rankings: An empirical framework. *Marketing Science 43*(3), 615–636.

Conlon, C. T. and J. H. Mortimer (2013). Demand estimation under incomplete product availability. *American Economic Journal: Microeconomics 5*(4), 1–30.

Crawford, G. S. and M. Shum (2005). Uncertainty and Learning in Pharmaceutical Demand. *Econometrica 73*(4), 1137–1173.

Davison, A. C. and D. V. Hinkley (1997). *Bootstrap methods and their application.* Number 1. Cambridge university press.

Erdem, T. and M. P. Keane (1996). Decision-Making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets. *Marketing Science 15*(1), 1–20.

Fox, J. T., K. i. Kim, S. P. Ryan, and P. Bajari (2012). The random coefficients logit model is identified. *Journal of Econometrics 166*(2), 204–212.

Gabaix, X. (2019). Behavioral inattention. In *Handbook of Behavioral Economics: Applications and Foundations 1*, Volume 2, pp. 261–343. Elsevier.

Gabaix, X., D. Laibson, G. Moloche, and S. Weinberg (2006). Costly information acquisition: Experimental analysis of a boundedly rational model. *American Economic Review 96*(4), 1043–1068.

Gardete, P. and M. Hunter (2020). Guiding consumers through lemons and peaches: An analysis of the effects of search design activities.

Gaynor, M., C. Propper, and S. Seiler (2016). Free to choose? reform, choice, and consideration sets in the english national health service. *American Economic Review*.

Goeree, M. S. (2008). Limited information and advertising in the us personal computer industry. *Econometrica 76*(5), 1017–1074.

Handel, B. R. and J. T. Kolstad (2015). Health insurance for "humans": Information frictions, plan choice, and consumer welfare. *American Economic Review 105*(8), 2449–2500.

Hastings, J. S. and L. Tejeda-Ashton (2008). Financial literacy, information, and demand elasticity: Survey and experimental evidence from mexico. Technical report, National Bureau of Economic Research.

Hastings, J. S. and J. M. Weinstein (2008). Information, school choice, and academic achievement: Evidence from two experiments. *The Quarterly journal of economics 123*(4), 1373–1414.

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

Honka, E. and P. Chintagunta (2017). Simultaneous or sequential? search strategies in the us auto insurance industry. *Marketing Science 36*(1), 21–42.

Honka, E., A. Hortaçsu, and M. A. Vitorino (2017). Advertising, consumer awareness, and choice: Evidence from the us banking industry. *RAND Journal of Economics 48*(3), 611–646.

Honka, E., A. Hortaçsu, and M. Wildenbeest (2019). Empirical search and consideration sets. In *Handbook of the Economics of Marketing*, Volume 1, pp. 193–257. Elsevier.

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. *Quarterly Journal of Economics 119*(2), 403–456.

Jindal, P. and A. Aribarg (2018). The value of price beliefs in consumer search.

Johnson, E. M. and M. M. Rehavi (2016). Physicians treating physicians: Information and incentives in childbirth. *American Economic Journal: Economic Policy 8*(1), 115–41.

Ke, T. T., Z.-J. M. Shen, and J. M. Villas-Boas (2016). Search for information on multiple products. *Management Science 62*(12), 3576–3603.

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

Kim, J. B., P. Albuquerque, and B. J. Bronnenberg (2017). The probit choice model under sequential search with an application to online retailing. *Management Science 63*(11), 3911–3929.

Kling, J. R., S. Mullainathan, E. Shafir, L. Vermeulen, and M. V. Wrobel (2008). Misperception in choosing medicare drug plans. *Unpublished manuscript*.

Manzini, P. and M. Mariotti (2014). Stochastic choice and consideration sets. *Econometrica 82*(3), 1153–1176.

Matzkin, R. L. (1993). Nonparametric identification and estimation of polychotomous choice models. *Journal of Econometrics 58*(1-2), 137–168.

McFadden, D. (1981). Econometric models of probabilistic choice. *Structural analysis of discrete data with econometric applications 198272*.

McFadden, D. L. (1984). Econometric analysis of qualitative response models. *Handbook of econometrics 2*, 1395–1457.

Mehta, N., S. Rajiv, and K. Srinivasan (2003). Price uncertainty and consumer search: A structural model of consideration set formation. *Marketing science 22*(1), 58–84.

Moraga-González, J. L., Z. Sándor, and M. R. Wildenbeest (2021). Consumer search and prices in the automobile market.

Roberts, J. H. and J. M. Lattin (1991). Development and testing of a model of consideration set composition. *Journal of Marketing Research*, 429–440.

Schwartz, B., A. Ward, J. Monterosso, S. Lyubomirsky, K. White, and D. R. Lehman (2002). Maximizing versus satisficing: Happiness is a matter of choice. *Journal of personality and social psychology 83*(5), 1178.

Simon, H. A. (1955). A behavioral model of rational choice. *The quarterly journal of economics*, 99–118.

Stahl, D. (1989). Oligopolistic pricing with sequential consumer search. *American Economic Review 79*(4), 700–712.

Ursu, R., S. Seiler, and E. Honka (2023). The sequential search model: A framework for empirical research.

Ursu, R. M. (2018). The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions. *Marketing Science 37*(4), 530–552.

Ursu, R. M., Q. Wang, and P. K. Chintagunta (2020). Search duration. *Marketing Science 39*(5), 849–871.

Weitzman, M. L. (1979). Optimal search for the best alternative. *Econometrica: Journal of the Econometric Society*, 641–654.

Wijsman, R. A. (1985). A useful inequality on ratios of integrals, with application to maximum likelihood estimation. *Journal of the American Statistical Association 80*(390), 472–475.

Woodward, S. E. and R. E. Hall (2012). Diagnosing consumer confusion and sub-optimal shopping effort: Theory and mortgage-market evidence. *American Economic Review 102*(7), 3249–76.

Appendix A: Additional Proofs

In this appendix, we collect the proofs not included in the main text. Throughout, we let $\mathcal{J} \equiv \{1, \ldots, J\}$ and often drop the i subscript for notational simplicity.

Here's a table with some of the notation used in the proofs.

Table 6: Notation

Notation	Definition	Appendix
P_4	$P(U_1 \geq U_k \; \forall k)$	A.1
P_5^S	$P\left(\{U_1 \geq U_k \forall k\} \cap \{EU_j \geq EU_1 \text{ for at least one } j \in \mathcal{S}\}\right.$ $\left.\cap \{g\left(x_1, \epsilon_1, U_j\right) \leq 0 \text{ for all } j \in \mathcal{S}\} \cap \{g\left(x_1, \epsilon_1, U_j\right) \geq 0 \text{ for all } j \in \mathcal{J}_{-1} \backslash \mathcal{S}\}\right)$	A.1
$P_{5,1}^S$	$P\left(\{U_1 \geq U_k \; \forall k\} \cap \{g\left(x_1, \epsilon_1, U_j\right) \leq 0 \text{ for all } j \in \mathcal{S}\} \cap \{g\left(x_1, \epsilon_1, U_j\right) \geq 0 \text{ for all } j \in \mathcal{J}_{-1} \backslash \mathcal{S}\}\right)$	A.1
$P_{5,2}^S$	$P(\{EU_1 \geq EU_j \text{ for all } j \in \mathcal{S}\} \cap \{g\left(x_1, \epsilon_1, U_j\right) \leq 0 \text{ for all } j \in \mathcal{S}\}$ $\cap \{g\left(x_1, \epsilon_1, U_j\right) \geq 0 \text{ for all } j \in \mathcal{J}_{-1} \backslash \mathcal{S}\})$	A.1
$P_{j,2}^*$	$P(\{U_j \geq U_{-j}\} \cap \{EU_{-j} \geq EU_j\} \cap \{g_i(x_j, \epsilon_j, U_{-j}) \leq 0\})$	A.2
$P_{j,3}^*$	$P(\{U_{-j} \geq U_j\} \cap \{EU_j \geq EU_{-j}\} \cap \{g_i(x_{-j}, \epsilon_{-j}, U_j) \leq 0\})$.	A.2
$P_{4new}, P_{5new}^S,$ $P_{5new,1}^S, P_{5new,2}^S$	Analog of $P_4, P_5^S, P_{5,1}^S, P_{5,2}^S$ with observables impacting search but not utility	A.5
$P_{1,sim}$	$P(U_1 \geq U_2)$	A.7
$P_{2,sim}$	$P(\{U_1 > U_2\} \cap \{EU_2 > EU_1\} \cap \{g_{sim}(EU_1 - EU_2) < 0\})$	A.7
$P_{4,out}, P_{5,out}^S,$ $P_{5,1,out}^S, P_{5,2,out}^S$	Analog of $P_4, P_5^S, P_{5,1}^S, P_{5,2}^S$ with outside option available without search	A.9
$P_{6,out}$	$P\left(\{U_1 \geq U_k \; \forall k \in \mathcal{J} \cup \{0\}\} \cap \{g(x_1, \epsilon_1, U_0) \leq 0\}\right)$	A.9

A.1 Proof of Lemma 2 when $J \geq 2$

In this section, we prove Lemma 2 for the more general case in which $J \geq 2$.

Proof. Let $\mathcal{J}_{-1} \equiv \{2, \ldots, J\}$, $\tilde{u}_j \equiv \alpha x_j + \beta z_j$ for all j, and $\tilde{\mathbf{u}} \equiv (\tilde{u}_1, \ldots, \tilde{u}_J)$. Similarly, we let $\tilde{eu}_j = \beta\gamma_0 + (\alpha + \beta\gamma_1)x_j$ and $\tilde{\mathbf{eu}} = (\tilde{eu}_1, \ldots, \tilde{eu}_J)$. Then, by (3) we can write for all $(\mathbf{x}, \mathbf{z})$ with $\tilde{z}_1 \geq \tilde{z}_j$

for all j:

$$
\begin{aligned}
s_1 = P(U_1 \geq U_k \ \forall k) - & \sum_{\mathcal{S} \subset \mathcal{J}_{-1}, \mathcal{S} \neq \emptyset} P\left(\{U_1 \geq U_k \forall k\} \cap \{EU_j \geq EU_1 \text{ for at least one } j \in \mathcal{S}\}\right. \\
& \left.\cap \{g(x_1, \epsilon_1, U_j) \leq 0 \text{ for all } j \in \mathcal{S}\} \cap \{g(x_1, \epsilon_1, U_j) \geq 0 \text{ for all } j \in \mathcal{J}_{-1} \backslash \mathcal{S}\}\right) \\
& \equiv P_4(\tilde{\mathbf{u}}) - \sum_{\mathcal{S} \subset \mathcal{J}_{-1}, \mathcal{S} \neq \emptyset} P_5^{\mathcal{S}}(\tilde{\mathbf{u}}, \tilde{\mathbf{eu}}, x_1)
\end{aligned} \tag{19}
$$

Further, for every $\mathcal{S} \subset \mathcal{J}_{-1}, \mathcal{S} \neq \emptyset$, we have

$$
\begin{aligned}
P_5^{\mathcal{S}} = & \ P\left(\{U_1 \geq U_k \ \forall k\} \cap \{g(x_1, \epsilon_1, U_j) \leq 0 \text{ for all } j \in \mathcal{S}\} \cap \{g(x_1, \epsilon_1, U_j) \geq 0 \text{ for all } j \in \mathcal{J}_{-1} \backslash \mathcal{S}\}\right) - \\
& P\left(\{U_1 \geq U_k \ \forall k\} \cap \{EU_1 \geq EU_j \text{ for all } j \in \mathcal{S}\}\right. \\
& \left.\cap \{g(x_1, \epsilon_1, U_j) \leq 0 \text{ for all } j \in \mathcal{S}\} \cap \{g(x_1, \epsilon_1, U_j) \geq 0 \text{ for all } j \in \mathcal{J}_{-1} \backslash \mathcal{S}\}\right) \\
= & \ P\left(\{U_1 \geq U_k \ \forall k\} \cap \{g(x_1, \epsilon_1, U_j) \leq 0 \text{ for all } j \in \mathcal{S}\} \cap \{g(x_1, \epsilon_1, U_j) \geq 0 \text{ for all } j \in \mathcal{J}_{-1} \backslash \mathcal{S}\}\right) - \\
& P\left(\{EU_1 \geq EU_j \text{ for all } j \in \mathcal{S}\} \cap \{g(x_1, \epsilon_1, U_j) \leq 0 \text{ for all } j \in \mathcal{S}\} \cap \{g(x_1, \epsilon_1, U_j) \geq 0 \text{ for all } j \in \mathcal{J}_{-1} \backslash \mathcal{S}\}\right) \\
\equiv & \ P_{5,1}^{\mathcal{S}}(\tilde{\mathbf{u}}, x_1) - P_{5,2}^{\mathcal{S}}(\tilde{\mathbf{u}}_{-1}, \tilde{\mathbf{eu}}, x_1)
\end{aligned}
$$

where $\tilde{\mathbf{u}}_{-1} \equiv (\tilde{u}_2, \ldots, \tilde{u}_J)$. The first equality follows from basic set algebra while the second follows from the fact that for all $j \in \mathcal{S}$ and all $k \in \mathcal{J}_{-1} \backslash \mathcal{S}$, (i) $EU_1 \geq EU_j$ implies $U_1 \geq U_j$ since $\tilde{z}_1 \geq \tilde{z}_j$ for all $j \in \mathcal{J}_{-1}$; and (ii) $g(x_1, \epsilon_1, U_k) \geq 0 \geq g(x_1, \epsilon_1, U_j)$ implies $U_k \leq U_j$, which (together with the implication in (i)) implies $U_1 \geq U_k$. Thus, the event $U_1 \geq U_k \ \forall k \in \mathcal{J}_{-1}$ is implied by the other events inside the probability and can be dropped. Now, note that $P_{5,2}^{\mathcal{S}}$ does not depend on z_1 and that $P_{5,1}^{\mathcal{S}}$ — as well as P_4 — only depends on x_j and z_j via $\tilde{u}_j$. Thus, the result follows from the chain rule. As in Lemma 2, identification of the distribution of ϵ is obtained by considering choice sets in which $\tilde{z}_k$ is the same for all k and varying $\mathbf{x}$. $\square$

A.2 Proof of Lemma 3

Let $P_{j,2}^*$ be the probability of failing to search — and thus choose — good j even if it is utility maximizing, and $P_{j,3}^*$ be the probability of choosing j even when it is not utility-maximizing (i.e., failing to search another, higher-utility good). Then, $s_j = P(U_j \geq U_k \forall k) - P_{j,2}^*(\tilde{\mathbf{u}}, \tilde{\mathbf{eu}}, \mathbf{x}) + P_{j,3}^*(\tilde{\mathbf{u}}, \tilde{\mathbf{eu}}, \mathbf{x})$, where $\tilde{u}_j = \alpha x_j + \beta z_j$, $\tilde{\mathbf{u}} \equiv (\tilde{u}_1, \ldots, \tilde{u}_J)$, $\tilde{eu}_j = \beta\gamma_0 + (\alpha + \beta\gamma_1)x_j$, $\tilde{\mathbf{eu}} \equiv (\tilde{eu}_1, \ldots, \tilde{eu}_J)$. Note that $P_{j,2}^*$ and $P_{j,3}^*$ depend on $\mathbf{x}$ via three channels: via the deterministic part of the utilities ($\tilde{\mathbf{u}}$), via the deterministic part of the of the expected utilities ($\tilde{\mathbf{eu}}$), and directly via the g_i function, i.e. the decision of whether or not to search the various products. Differentiating, we obtain:

$$
\begin{aligned}
\frac{\partial s_j}{\partial z_j} &= \beta\left[\frac{\partial P(U_j \geq U_k \forall k)}{\partial \tilde{u}_j} - \frac{\partial P^*_{j,2}}{\partial \tilde{u}_j}(\tilde{\mathbf{u}}, \tilde{\mathbf{eu}}, \mathbf{x}) + \frac{\partial P^*_{j,3}}{\partial \tilde{u}_j}(\tilde{\mathbf{u}}, \tilde{\mathbf{eu}}, \mathbf{x})\right] \\
\frac{\partial s_j}{\partial x_j} &= \alpha\left[\frac{\partial P(U_j \geq U_k \forall k)}{\partial \tilde{u}_j} - \frac{\partial P^*_{j,2}}{\partial \tilde{u}_j}(\tilde{\mathbf{u}}, \tilde{\mathbf{eu}}, \mathbf{x}) + \frac{\partial P^*_{j,3}}{\partial \tilde{u}_j}(\tilde{\mathbf{u}}, \tilde{\mathbf{eu}}, \mathbf{x})\right] - \frac{\partial P^*_{j,2}}{\partial x_j}(\tilde{\mathbf{u}}, \tilde{\mathbf{eu}}, \mathbf{x}) + \frac{\partial P^*_{j,3}}{\partial x_j}(\tilde{\mathbf{u}}, \tilde{\mathbf{eu}}, \mathbf{x}) \\
&\quad -(\alpha + \beta\gamma_1)\frac{\partial P^*_{j,2}}{\partial \tilde{eu}_j}(\tilde{\mathbf{u}}, \tilde{\mathbf{eu}}, \mathbf{x}) + (\alpha + \beta\gamma_1)\frac{\partial P^*_{j,3}}{\partial \tilde{eu}_j}(\tilde{\mathbf{u}}, \tilde{\mathbf{eu}}, \mathbf{x}).
\end{aligned}
$$

Note that $\frac{\partial P(U_j \geq U_{-j})}{\partial \tilde{u}_j} - \frac{\partial P^*_{j,2}}{\partial \tilde{u}_j}(\tilde{\mathbf{u}}, \tilde{\mathbf{eu}}, \mathbf{x}) + \frac{\partial P^*_{j,3}}{\partial \tilde{u}_j}(\tilde{\mathbf{u}}, \tilde{\mathbf{eu}}, \mathbf{x}) = \frac{\partial s_j}{\partial \tilde{u}_j} \geq 0$.[32] Further, $\frac{\partial P^*_{j,2}}{\partial x_j}(\tilde{\mathbf{u}}, \tilde{\mathbf{eu}}, \mathbf{x}) \leq 0$ and $\frac{\partial P^*_{j,3}}{\partial x_j}(\tilde{\mathbf{u}}, \tilde{\mathbf{eu}}, \mathbf{x}) \geq 0$ due to our assumptions about the function g. Finally, $\frac{\partial P^*_{j,2}}{\partial \tilde{eu}_j}(\tilde{\mathbf{u}}, \tilde{\mathbf{eu}}, \mathbf{x}) \leq 0$ and $\frac{\partial P^*_{j,3}}{\partial \tilde{eu}_j}(\tilde{\mathbf{u}}, \tilde{\mathbf{eu}}, \mathbf{x}) \geq 0$ by Assumption 2(*i*). Therefore, normalizing $\alpha = 1$, we obtain $\left|\frac{\partial s_j}{\partial z_j} / \frac{\partial s_j}{\partial x_j}\right| \leq \beta$ under the assumption that $\alpha + \beta\gamma_1 \geq 0$. One can show that $\left|\frac{\partial s_j}{\partial z_k} / \frac{\partial s_j}{\partial x_k}\right| \leq \beta$ for $k \neq j$ using an analogous argument.

A.3 Identifying good 1 when z_j is vector-valued in the linear homogeneous case

For simplicity, the results in the main text are for the case where z_j is scalar-valued for all goods j. This implies that one can label good 1 as the good with the highest value of $\tilde{z}$ without loss of generality. As we have noted, if there are multiple z attributes per good, then our results apply if the data contains one choice set where one good is preferable to all other goods on each of the $\tilde{z}$ attributes. This is not without loss.

The basic idea of our approach is that one can define a weighted average of the relevant $z's$ where the weights are recoverable from first derivatives. Then, we proceed as usual, defining good 1 as the good with the highest value of this weighted average. More formally, let z_{kj} be the $k-$th hidden attribute of good j and let β_k be the associated preference parameter. As above, let $\tilde{u}_j = \alpha x_j + \beta z_j$. By Assumption 2, we can write $s_j = f_{s_j}(\tilde{u}_1, \ldots, \tilde{u}_J, x_1, \ldots, x_J)$ for all j and thus $\frac{\partial s_j}{\partial z_{kj}} = \frac{\partial f_{s_j}}{\partial \tilde{u}_j}\beta_k$, implying $\frac{\partial s_j}{\partial z_{kj}} / \frac{\partial s_j}{\partial z_{k'j}} = \beta_k / \beta_{k'}$ for all k, k'. This means that we can compare the hidden component of utility across goods. Specifically, letting $\beta_1 > 0$ without loss, we have that, for any pair of goods j and j', $\sum_k \beta_k \tilde{z}_{kj} \geq \sum_k \beta_k \tilde{z}_{kj'}$ if and only if $\tilde{z}_{1j} - \tilde{z}_{1j'} + \sum_{k>1} \frac{\beta_k}{\beta_1}\left(\tilde{z}_{kj} - \tilde{z}_{kj'}\right) \geq 0$. Since the left-hand side of the last inequality is identified, we can rank goods based on the component of utility that is not known pre-search. Lemma 2 then applies by defining good 1 as the good with the highest value of $\sum_k \beta_k \tilde{z}_{kj}$. Note that such a good always exists in any choice set (excluding ties) since $\sum_k \beta_k \tilde{z}_{kj}$ is scalar-valued.

[32] Increasing $\tilde{u}_j$ can only switch consumers from not choosing good j to choosing j but never the reverse. To see this, note first that conditional on searching any given set of goods, increasing $\tilde{u}_j$ increases the probability that good j is chosen. Second, changing $\tilde{u}_j$ doesn't change the probability that good j is searched, which depends on $g_i(x_j, \epsilon_j, U_{-j})$ for each alternative searched good. Third, changing $\tilde{u}_j$ never makes other goods more likely to be searched. Specifically, an alternative good k is searched if and only if $g_i(x_k, \epsilon_k, U_{k'}) \geq 0$ for all goods k' currently searched. This quantity is unchanged for $k' \neq j$ and weakly decreasing for $k' = j$, so no good can become more likely to be searched. Therefore, $\frac{\partial s_j}{\partial \tilde{u}_j} \geq 0$.

A.4 Endogenous attributes

Here, we show how to extend our results to the case where some product attributes are endogenous (Section 3.1). Consistent with standard results on nonparametric identification of demand with endogeneity (Berry and Haile (2014)), we assume the data contains information about choice probabilities across many markets. This is in contrast to the case without endogeneity where just six markets are in principle sufficient. We start by considering the case where price is visible prior to search. Letting $\delta = (\delta_1, \ldots, \delta_J)$ where δ_j is defined in Assumption 3(i) and $\tilde{z}_j = z_j - (\gamma_0 + \gamma_1 x_j + \gamma_{1,p} p_j)$, we may write the choice probability of good j as

$$s_j = \sigma_j (\delta, \tilde{\mathbf{z}}) \tag{20}$$

for some function σ_j. Repeating this for all j and stacking the equations, we obtain a demand system of the form

$$\mathbf{s} = \sigma (\delta, \tilde{\mathbf{z}}) \tag{21}$$

where $\mathbf{s} = (s_1, \ldots, s_J)$. We also define the choice probability of the outside option as $s_0 \equiv 1 - \sum_{j=1}^{J} s_j$, with associated function $\sigma_0 (\delta, \tilde{\mathbf{z}})$. We establish nonparametric identification of this demand system by invoking results from Berry and Haile (2014) (henceforth, BH).[33] Specifically, the results in BH yield identification of $(\xi_j)_{j=1}^{J}$ for every market in the population. This means that all the arguments of σ are known, which immediately implies (nonparametric) identification of σ itself. Once σ is identified, one may apply our results in Section 2.2 to identify the preference parameters α, β and λ. Note that, while knowledge of σ is sufficient for several counterfactuals of interest (e.g., computing equilibrium prices after a potential merger or tax), the preference parameters are required to predict how choices and welfare would change if consumers were given full information, among other things. In this sense, our approach complements the identification results in BH within the class of search models we consider.

To prove identification of σ, we first note that model (20) satisfies the index restriction in BH's Assumption 1. Second, we assume that we have excluded instruments $\mathbf{w}$ which, together with the exogenous attributes, satisfy the following mean-independence restriction

$$E (\xi_j | \mathbf{x}, \tilde{\mathbf{z}}, \mathbf{w}) = 0 \quad \text{for all } j \tag{22}$$

almost surely (Assumption 3 in BH) and assume that the instruments shift the endogenous variables (choice probabilities and endogenous prices $\mathbf{p}$) to a sufficient degree (as in BH's Assumption 4). The endogeneity problem is exactly as in the full information case; thus, all of the instruments that are commonly used (e.g., cost shifters, exogenous attributes of competing products, Hausman instruments)

[33] See also Berry, Gandhi, and Haile (2013).

can be invoked in our context. Finally, we verify that the demand system satisfies the "connected substitutes" restriction defined in BH's Assumption 2. To this end, we prove the following result.

Lemma 5. *Let utility be given by* (10) *with* ϵ_i *supported on* $\mathbb{R}^J$ *and let Assumptions 2(i), 2(iii), 2(iv), and 3(i) hold. Then, for all* $j, k = 1, \ldots, J$ *with* $j \neq k$, σ_j *is (i) strictly increasing in* δ_j *and (ii) strictly decreasing in* δ_k.

Proof. Fix (δ_j, z_j) for all j. To prove claim (i), we show that an increase in δ_j can only induce a consumer to switch from not choosing j to choosing j but never vice versa, and that a positive mass of consumers will switch to choosing j. To see this, consider the case where consumer i initially searches j, which happens if and only if $g_i(\delta_j, \epsilon_{ij}, U_{ik}) \geq 0$ for all k such that $EU_{ik} \geq EU_{ij}$. Let $\Delta \geq 0$ be the change in δ_j. Since g_i is increasing in its first argument, we have $g_i(\delta_j + \Delta, \epsilon_{ij}, U_{ik}) \geq 0$ for all k such that $EU_{ik} \geq EU_{ij} + \Delta$ and thus i will still search j. Moreover, since g_i is decreasing in its last argument, if $g_i(\delta_k, \epsilon_{ik}, U_{ij}) \leq 0$ for some k such that $EU_{ik} \leq EU_{ij}$ (i.e., if k is initially not searched), then $g_i(\delta_k, \epsilon_{ik}, U_{ij} + \Delta) \leq 0$ (i.e., k is also not searched after the change in δ_j), which means that the set of goods searched by i never becomes larger. Next, note that if $U_{ij} \geq U_{ik}$ for all k in the set of searched goods $\mathcal{G}_i$, then $U_{ij} + \Delta \geq U_{ik}$ for all $k \in \mathcal{G}_i$. Further, since ϵ_i is supported on all of $\mathbb{R}^J$, there is a positive mass of consumers for which $U_{ik} \geq U_{ij}$ for some $k \in \mathcal{G}_i$, but $U_{ij} + \Delta \geq U_{ik}$ for all $k \in \mathcal{G}_i$. An analogous argument proves claim (ii). $\square$

Lemma 5 implies that the goods are connected substitutes in δ (see Definition 1 in BH), which in turn allows us to prove identification of σ by invoking Theorem 1 in BH.[34] Specifically, we can invert the demand system σ for the indices δ and write

$$\beta\gamma_0 + (\alpha + \beta\gamma_1) x_j + (\lambda + \beta\gamma_{1,p}) p_j + \xi_j = \sigma_j^{-1}(\mathbf{s}, \tilde{\mathbf{z}}) \tag{23}$$

for all j. Equations (22) and (23) naturally lead to a nonparametric instrumental variable approach to pin down σ_j^{-1} (and thus σ_j).[35]

Nothing in the argument above hinges on the fact that prices are visible to consumers prior to search. In particular, when consumers need to search to reveal prices, $z_j = p_j$ and equation (21) becomes

$$\mathbf{s} = \sigma(\underline{\delta}, \tilde{\mathbf{p}}),$$

where $\underline{\delta}_j$ is defined in Assumption 3(*ii*) and $\tilde{\mathbf{p}}$ denotes the vector of the components of prices not expected by consumers prior to search. Since the proof of Lemma 5 continues to apply if Assumption

[34] Note that the proof of Theorem 1 in BH only uses the fact that goods are connected substitutes in δ, not in $-\mathbf{p}$.

[35] Compiani (2022) proposes to approximate σ_j^{-1} using Bernstein polynomials. We use a similar approach in Section 5 to estimate the demand function for the case without endogeneity.

3(i) is replaced with 3(ii), one can follow the argument above to identify the function σ when prices are only revealed via search.

A.5 Identification when Observables Impact Search but not Utility

Here, we state and prove the results described in Section 3.2. We make the following assumptions:

Assumption 4. (i) If consumer i searches j, then i also searches all j' s.t. $m\left(EU_{ij'}, r_{j'}\right) \geq m\left(EU_{ij}, r_j\right)$, where m is strictly increasing in both arguments;

(ii) There is at least one good $j \neq 1$ such that $r_j > r_1$;

(iii) The support of $(\mathbf{x}, \mathbf{z})\Big|\mathbf{r}$ has positive Lebesgue measure for all $\mathbf{r} \equiv (r_1, \ldots, r_J)$.

(iv) The search model admits a discrete choice representation that also satisfies the independence of irrelevant alternatives (IIA) property.

Assumption 4(iii) means that, for identification purposes, we consider variation in product characteristics holding fixed product search position. In practice, search position is likely to vary as a function of observables (e.g., products are sorted in order of price). However, because of the discrete nature of search position, we are likely to see variation conditional on search position and this is the variation we will use to identify our model. Assumption 4(iv) requires that consumers' search behavior can be represented as a standard discrete choice model satisfying IIA. As shown in Armstrong (2017), the Weitzman (1979) sequential search model (see Example 1) can be represented as a discrete choice model where consumers maximize product-specific indices defined as the minimum between the utility and the reservation value for each product. Then, Assumption 4(iv) is satisfied by letting the ϵ_{ij} be Gumbel distributed.

If the order in which consumers search does not just depend on the expected utilities, but on the variable r as well, Lemma 1 will no longer hold as stated: the good with the highest value of $\tilde{z}_j$ can be searched, another good j' may have higher utility (and thus higher expected utility), but good j' may not be searched because it has lower search position. However, an extension of Lemma 1 will still hold in this case, which then allows us to prove identification of preferences.

Lemma 6. *Let Assumptions 1, 2(ii)-2(iv), and 4 hold. Then, if consumer i searches good 1 (i.e., the good with the highest value of $\tilde{z}$), then i chooses the good which maximizes utility among all goods with $r_j \geq r_1$.*

Proof. Suppose there was a good j with $r_j > r_1$ and $U_{ij} > U_{i1}$ that consumer i does not search. We can follow the proof of Lemma 1 to show that $EU_{ij} > EU_{i1}$. By Assumption 4(i), this implies that good j is searched, which is a contradiction. □

In words, if higher search position only makes a good more likely to be searched, then goods with higher utility *and* higher search position will always be searched if good 1 is searched. Given this lemma, we can apply a modification of the identification argument in Lemma 2 after conditioning on the subset of goods with higher search position than good 1. As in Lemma 2, we impose a location normalization ($\epsilon_{i\tilde{j}} = 0$ for some $\tilde{j}$ and all i) and a scale normalization ($\alpha = 1$).

Lemma 7. *Let the assumptions of Lemma 6 hold. Let $s_{1|\mathcal{R}}$ denote the choice probability for good 1 conditional on consumers choosing in $\mathcal{R} = \{j : r_j \geq r_1\}$ and assume that $\frac{\partial^2 s_{1|\mathcal{R}}}{\partial z_1 \partial x_{j^*}}(\mathbf{x}^*, \mathbf{z}^*, \mathbf{r}^*) \neq 0$ for some $(\mathbf{x}^*, \mathbf{z}^*, \mathbf{r}^*)$ and $j^* \neq 1$ and that $s_{1|\mathcal{R}}$ is twice differentiable. Then, β is identified. Additionally, the distribution of ϵ is nonparametrically identified if the support of $\left(\epsilon_k - \epsilon_{\tilde{j}}\right)_{k \neq \tilde{j}}$ is a subset of $\left\{(\alpha + \beta\gamma_1)\left(x_{\tilde{j}} - x_k\right)_{k \neq \tilde{j}} : \gamma_1 (x_{\tilde{j}} - x_k)_{k \neq \tilde{j}} = \left(z_{\tilde{j}} - z_k\right)_{k \neq \tilde{j}} \text{ for some } (\mathbf{x}, \mathbf{z}) \text{ in its support}\right\}$ and the supports of $\tilde{\mathbf{z}}$ and $\mathbf{r}$ contain a point such that $\tilde{z}_j = \tilde{z}_k$ and $r_j = r_k$ for all j, k, respectively.*

Proof. Under Assumption 4(*iv*), the choice probability for good 1 conditional on consumers choosing in $\mathcal{R}$, $s_{1|\mathcal{R}}$, is equal to the choice probability for good 1 if consumers only faced $\mathcal{R}$ as their choice set. Further, by Lemma 6, the only mistake a consumer can make when faced with choice set $\mathcal{R}$ is to fail to search good 1 when it is in fact the good with the highest utility in $\mathcal{R}$. Thus, letting $\mathcal{R}_{-1} = \mathcal{R} \setminus \{1\}$, we can write for all $(\mathbf{x}, \mathbf{z})$ with $\tilde{z}_1 \geq \tilde{z}_j$ for all j:

$$\begin{aligned} s_{1|\mathcal{R}} = P(U_1 \geq U_k \ \forall k \in \mathcal{R}) - \sum_{\mathcal{S} \subset \mathcal{R}_{-1}, \mathcal{S} \neq \emptyset} & P\left(\{U_1 \geq U_k \forall k \in \mathcal{R}\} \cap \{m(EU_j, r_j) \geq m(EU_1, r_1) \text{ for some } j \in \mathcal{S}\}\right. \\ & \left. \cap \{g(x_1, \epsilon_1, U_j) \leq 0 \text{ for all } j \in \mathcal{S}\} \cap \{g(x_1, \epsilon_1, U_j) \geq 0 \text{ for all } j \in \mathcal{R}_{-1} \backslash \mathcal{S}\}\right) \\ \equiv \ P_{4new}(\tilde{\mathbf{u}}) - \sum_{\mathcal{S} \subset \mathcal{R}_{-1}, \mathcal{S} \neq \emptyset} & P_{5new}^{\mathcal{S}}(\tilde{\mathbf{u}}, \tilde{\mathbf{eu}}, x_1, \mathbf{r}). \end{aligned} \tag{24}$$

Further, for every $\mathcal{S} \subset \mathcal{R}_{-1}, \mathcal{S} \neq \emptyset$, we have

$$\begin{aligned} P_{5new}^{\mathcal{S}} = \ & P\left(\{U_1 \geq U_k \ \forall k \in \mathcal{R}\} \cap \{g(x_1, \epsilon_1, U_j) \leq 0 \text{ for all } j \in \mathcal{S}\} \cap \{g(x_1, \epsilon_1, U_j) \geq 0 \text{ for all } j \in \mathcal{R}_{-1} \backslash \mathcal{S}\}\right) - \\ & P\left(\{U_1 \geq U_k \ \forall k \in \mathcal{R}\} \cap \{m(EU_1, r_1) \geq m(EU_j, r_j) \text{ for all } j \in \mathcal{S}\}\right. \\ & \left. \cap \{g(x_1, \epsilon_1, U_j) \leq 0 \text{ for all } j \in \mathcal{S}\} \cap \{g(x_1, \epsilon_1, U_j) \geq 0 \text{ for all } j \in \mathcal{R}_{-1} \backslash \mathcal{S}\}\right) \\ = \ & P\left(\{U_1 \geq U_k \ \forall k \in \mathcal{R}\} \cap \{g(x_1, \epsilon_1, U_j) \leq 0 \text{ for all } j \in \mathcal{S}\} \cap \{g(x_1, \epsilon_1, U_j) \geq 0 \text{ for all } j \in \mathcal{R}_{-1} \backslash \mathcal{S}\}\right) - \\ & P\left(\{m(EU_1, r_1) \geq m(EU_j, r_j) \ \forall j \in \mathcal{S}\} \cap \{g(x_1, \epsilon_1, U_j) \leq 0 \ \forall j \in \mathcal{S}\} \cap \{g(x_1, \epsilon_1, U_j) \geq 0 \ \forall j \in \mathcal{R}_{-1} \backslash \mathcal{S}\}\right) \\ \equiv \ & P_{5new,1}^{\mathcal{S}}(\tilde{\mathbf{u}}, x_1, \mathbf{r}) - P_{5new,2}^{\mathcal{S}}(\tilde{\mathbf{u}}_{-1}, \tilde{\mathbf{eu}}, x_1, \mathbf{r}), \end{aligned}$$

where $\tilde{\mathbf{u}}_{-1} \equiv (\tilde{u}_2, \ldots, \tilde{u}_J)$. The first equality follows from basic set algebra while the second follows from the fact that the event $\{U_1 \geq U_j \ \ \forall j \in \mathcal{R}\}$ is implied by the other events inside the second probability, since (i) if $j \in \mathcal{S}$, then $m(EU_1, r_1) \geq m(EU_j, r_j)$ implies $EU_1 \geq EU_j$, which in turn implies $U_1 \geq U_j$; (ii) if $j \notin \mathcal{S}$, then $g(x_1, \epsilon_1, U_j) \geq 0 \geq g(x_1, \epsilon_1, U_k)$ for all $k \in \mathcal{S}$ implies $U_j \leq U_k$. Note that $P_{5new,2}^{\mathcal{S}}$ does not depend on z_1 and $P_{5new,1}^{\mathcal{S}}(\tilde{\mathbf{u}}, x_1, \mathbf{r})$ only depends on x_j and z_j via $\tilde{u}_j$ for $j \neq 1$, so that $\frac{\partial^2 s_{1|\mathcal{R}}}{\partial z_1 \partial z_{j^*}}(\mathbf{x}^*, \mathbf{z}^*, \mathbf{r}^*) \Big/ \frac{\partial^2 s_{1|\mathcal{R}}}{\partial z_1 \partial x_{j^*}}(\mathbf{x}^*, \mathbf{z}^*, \mathbf{r}^*) = \frac{\beta}{\alpha}$. Note that $s_{1|\mathcal{R}}$ can be estimated by taking $\mathcal{R}$ as the choice set faced by consumers and dropping those consumers that choose products outside $\mathcal{R}$. Thus, $\frac{\beta}{\alpha}$ is identified, and so is β given the normalization $\alpha = 1$. Finally, we identify the distribution of ϵ by looking at choice sets in which $\tilde{z}_j = \tilde{z}$ and $r_j = r$ for all j. Consumers facing such choice sets always choose the utility-maximizing good, since the first good they search — the one that maximizes $m(EU_j, r_j)$ — also maximizes utility. Thus, by varying $\mathbf{x}$, we can identify the distribution of ϵ just like in Lemma 2. Note that if the support of $\mathbf{r}$ does not contain points where $r_j = r$ for all j (which would be the case if, e.g., r is a product's rank on the webpage), this last step requires extrapolation outside the support of $\mathbf{r}$ and thus cannot be done nonparametrically. Further, even when this is not an issue, restricting attention to the slice of the data around $r_j = r$ for all r is likely to lead to slower-than-parametric convergence rates.

□

A.6 Unobservables revealed by search

Here, we show that the ratio of second derivatives in (4) robustly identifies $\frac{\beta}{\alpha}$ in the model where ϵ_{ij} is revealed to consumer i only upon searching good j (Section 3.3). Order goods in increasing order of x. Then, if $\alpha + \beta\gamma_1 \geq 0$, consumers search in descending order of x. Thus, for $j = 1, \ldots, J$,

$$\begin{aligned}
s_j &= \sum_{k=1}^{j} P\left(\{U_j \geq U_{j'}\ \forall\, j' \in \{k, \ldots, J\}\} \cap \{\text{search exactly } k, \ldots, J\}\right) \\
&= \sum_{k=1}^{j} P(\{U_j \geq U_{j'}\ \forall\, j' \in \{k, \ldots, J\}\} \cap \{g(x_h, U_{h'}) \geq 0\ \forall\, h = k, \ldots, J-1; h' \in \{h+1, \ldots, J\}\} \cap \\
&\quad \{g(x_h, U_j) \leq 0\ \forall h = 1, \ldots, k-1\}) \\
&\equiv \sum_{k=1}^{j} P_j^{(k)}\left(\tilde{\mathbf{u}}, \mathbf{x}_{-J}\right),
\end{aligned}$$

where $\tilde{u}_j = \alpha x_j + \beta z_j$ and $\tilde{\mathbf{u}} = (\tilde{u}_1, \ldots, \tilde{u}_J)$, as above, and $\mathbf{x}_{-J} = (x_1, \ldots, x_{J-1})$. Thus,

$$\begin{aligned}
\frac{\partial^2 s_j}{\partial z_j \partial z_J} &= \sum_{k=1}^{j} \frac{\partial^2 P_j^{(k)}}{\partial \tilde{u}_j \partial \tilde{u}_J} \beta^2 \\
\frac{\partial^2 s_j}{\partial z_j \partial x_J} &= \sum_{k=1}^{j} \frac{\partial^2 P_j^{(k)}}{\partial \tilde{u}_j \partial \tilde{u}_J} \alpha\beta
\end{aligned}$$

Thus, the ratio of the latter two derivatives identifies $\frac{\beta}{\alpha}$. Note that the ratio of $\frac{\partial s_j}{\partial z_J}$ to $\frac{\partial s_j}{\partial x_J}$ for any j would also work, thus providing a test for the null hypothesis that ϵ is only revealed via search. On the other hand, for $j < J$,

$$\begin{aligned}
\frac{\partial s_j}{\partial z_j} &= \sum_{k=1}^{j} \frac{\partial P_j^{(k)}}{\partial \tilde{u}_j} \beta \\
\frac{\partial s_j}{\partial x_j} &= \sum_{k=1}^{j} \left(\frac{\partial P_j^{(k)}}{\partial \tilde{u}_j} + \frac{1}{\alpha} \frac{\partial P_j^{(k)}}{\partial x_j} \right) \alpha
\end{aligned} \tag{25}$$

Since $\frac{1}{\alpha}\frac{\partial P_j^{(k)}}{\partial x_j} \geq 0$, (25) implies that the ratio of first derivatives suffers from attenuation bias, i.e. $\frac{\frac{\partial s_j}{\partial z_j}}{\frac{\partial s_j}{\partial x_j}} \leq \frac{\beta}{\alpha}$.

A.7 Simultaneous search à la Honka et al. (2017)

While our main model allows for consumers to choose which goods to search in one simultaneous step, one form of simultaneous search that is not accommodated is that in which a consumer optimally chooses the number K of goods to uncover and then proceeds to simultaneously search the top K in terms of expected utility (see Chade and Smith (2006) and, for an application, Honka, Hortaçsu, and Vitorino (2017)). Our framework from Section 2 does not subsume this model since in this case the decision of whether or not to search good j depends not only on the expected utility of good j, but on the expected utility of all other goods as well, thus violating Assumption 2(ii).

In this appendix, we show via an alternative argument that the usual second-derivative ratio from equation (4) still identifies $\frac{\beta}{\alpha}$ in the two-good $K-$rank model. This complements the simulation results from Appendix D showing that our method succeeds in a model where consumers search the top K goods (with K varying randomly across consumers).

Consider the simultaneous search model in Honka, Hortaçsu, and Vitorino (2017) with $J = 2$ goods. In this model, a consumer looks at the expected utilities and decides whether to only search the good with the highest expected utility or search both goods, which entails a cost c. Consumers form expectations over the distribution of $(\tilde{z}_1, \tilde{z}_2)$, assumed to be independent of everything else. As usual, we denote by 1 the good with the highest value of $\tilde{z}$.

Note that consumer i searches 2 but not 1 if and only if $EU_{i2} > EU_{i1}$ and

$$E_{\tilde{z}_1,\tilde{z}_2}\left[\max\left\{EU_{i1} + \beta\tilde{z}_1, EU_{i2} + \beta\tilde{z}_2\right\}\right] - c < E_{\tilde{z}_2}\left[EU_{i2} + \beta\tilde{z}_2\right],$$

i.e.

$$E_{\tilde{z}_1,\tilde{z}_2}\left[\max\left\{EU_{i1} - EU_{i2} + \beta\left(\tilde{z}_1 - \tilde{z}_2\right), 0\right\}\right] - c < 0$$

or $g_{sim}\left(EU_{i1} - EU_{i2}\right) < 0$ for an increasing function g_{sim}. Equation (5) then can be written as

$$\begin{aligned} s_1 &= P(U_1 > U_2) - P(\{U_1 > U_2\} \cap \{EU_2 > EU_1\} \cap \{g_{sim}(EU_1 - EU_2) < 0\}) \\ &= P_{1,sim} - P_{2,sim} \end{aligned}$$

Letting $\tilde{u}_j = \alpha x_j + \beta z_j$ and $\tilde{eu}_j = \beta\gamma_0 + (\alpha + \beta\gamma_1)x_j$, we also have:

$$\frac{\partial^2 s_1}{\partial z_1 \partial z_2} = \beta^2\left(\frac{\partial^2 P_{1,sim}}{\partial\tilde{u}_1\partial\tilde{u}_2} - \frac{\partial^2 P_{2,sim}}{\partial\tilde{u}_1\partial\tilde{u}_2}\right) \tag{26}$$

and

$$\frac{\partial^2 s_1}{\partial z_1 \partial x_2} = \alpha\beta\left(\frac{\partial^2 P_{1,sim}}{\partial \tilde{u}_1 \partial \tilde{u}_2} - \frac{\partial^2 P_{2,sim}}{\partial \tilde{u}_1 \partial \tilde{u}_2}\right) - (\alpha + \beta\gamma_1)\beta \frac{\partial^2 P_{2,sim}}{\partial \tilde{u}_1 \partial \tilde{eu}_2} \tag{27}$$

So, if $\frac{\partial^2 P_{2,sim}}{\partial \tilde{u}_1 \partial \tilde{eu}_2} = 0$, then the ratio of (26) to (27) identifies $\frac{\beta}{\alpha}$. Note that the event in $P_{2,sim}$ is equivalent to the following set of inequalities: (i) $\epsilon_{i1} > \tilde{u}_2 - \tilde{u}_1 + \epsilon_{i2}$, (ii) $\epsilon_{i1} < \tilde{eu}_2 - \tilde{eu}_1 + \epsilon_{i2}$, (iii) $\epsilon_{i1} < g_{sim}^{-1}(0) + \tilde{eu}_2 - \tilde{eu}_1 + \epsilon_{i2}$, Then, letting $\tilde{\epsilon} = \epsilon_1 - \epsilon_2$, we have:

$$P_{2,sim} = \int_{\tilde{u}_2 - \tilde{u}_1}^{\min(\tilde{eu}_2 - \tilde{eu}_1, g_{sim}^{-1}(0) + \tilde{eu}_2 - \tilde{eu}_1)} f_{\tilde{\epsilon}}(\tilde{\epsilon}) d\tilde{\epsilon} \tag{28}$$

By Leibniz's rule, $\frac{\partial P_{2,sim}}{\partial \tilde{u}_1} = f_{\tilde{\epsilon}}(\tilde{u}_2 - \tilde{u}_1)$ and thus $\frac{\partial^2 P_{2,sim}}{\partial \tilde{u}_1 \partial \tilde{eu}_2} = 0$.

Finally, we show that the ratio of first derivatives leads to attenuation bias if $\alpha + \beta\gamma_1 \geq 0$. This follows directly from

$$\begin{aligned}
\frac{\partial s_1}{\partial z_1} &= \beta\left(\frac{\partial P_{1,sim}}{\partial \tilde{u}_1} - \frac{\partial P_{2,sim}}{\partial \tilde{u}_1}\right) \\
\frac{\partial s_1}{\partial x_1} &= \alpha\left(\frac{\partial P_{1,sim}}{\partial \tilde{u}_1} - \frac{\partial P_{2,sim}}{\partial \tilde{u}_1} - \frac{\alpha + \beta\gamma_1}{\alpha}\frac{\partial P_{2,sim}}{\partial \tilde{eu}_1}\right)
\end{aligned}$$

and the fact that $\frac{\partial P_{2,sim}}{\partial \tilde{eu}_1} < 0$.

A.8 Discrete attributes

Our main result applies to the case where one can take derivatives of choice probabilities with respect to the attribute of various goods, implying that the attributes must be a continuously distributed. Here, we show that our argument can be extended to the case where attributes are discrete provided derivatives are replaced with differences and at least one attribute is continuous. We begin with the the case where z is discrete and there is at least one continuous x attribute (additional discrete x attributes are allowed).

Let 1 denote one of the goods with the highest value of $\tilde{z}$ in the choice set. When z is discrete, there can be multiple such goods, but the argument holds for any one of them. Take any data point $(\mathbf{x}, \mathbf{z})$ and let $D_{\delta_{z_1}, z_1} s_1(\mathbf{x}, \mathbf{z}) \equiv s_1(\mathbf{x}, \mathbf{z} + \delta_{z_1}) - s_1(\mathbf{x}, \mathbf{z})$, where $\mathbf{z} + \delta_{z_1}$ is equal to $\mathbf{z}$ except that z_1 is increased or decreased by δ_{z_1}. By the same argument as in the proof of Lemma 2, $D_{\delta_{z_1}, z_1} s_1(\mathbf{x}, \mathbf{z})$ depends on x_j and z_j for $j \neq 1$ only via $\alpha x_j + \beta z_j$ (this relies on Lemma 1, which holds regardless of whether z is discrete or continuous). Thus, as in the continuous case, the marginal rate of substitution $\frac{\beta}{\alpha}$ will be identified by looking at how $D_{\delta_{z_1}, z_1} s_1(\mathbf{x}, \mathbf{z})$ changes as z_j and x_j vary for $j \neq 1$. Specifically, take a change from z_j to $z_j + \delta_{z_j}$ and a change from x_j to $x_j + \delta_{x_j}$ that lead to the same change in $\tilde{u}_j \equiv \alpha x_j + \beta z_j$. It must be that $\beta\delta_{z_j} = \alpha\delta_{x_j}$, implying that $\frac{\beta}{\alpha} = \frac{\delta_{x_j}}{\delta_{z_j}}$. Thus, given a discrete change δ_{z_j}, finding the corresponding change δ_{x_j} identifies the marginal rate of substitution. To this end, we look at how $D_{\delta_{z_1},z_1} s_1(\mathbf{x},\mathbf{z})$ changes as x_j moves continuously. The change δ_{x_j} is a solution of the following equation in d_{x_j}

$$D_{\delta_{z_1},z_1} s_1(\mathbf{x}+d_{x_j},\mathbf{z}) = D_{\delta_{z_1},z_1} s_1(\mathbf{x},\mathbf{z}+\delta_{z_j}), \tag{29}$$

where $\mathbf{z}+\delta_{z_j}$ is equal to $\mathbf{z}$ except that z_j is changed to $z_j+\delta_{z_j}$, and similarly for $\mathbf{x}+d_{x_j}$. If $D_{\delta_{z_1},z_1} s_1(\mathbf{x},\mathbf{z})$ is monotonic in $\tilde{u}_j$, δ_{x_j} is the unique solution of (29). More generally, we show in the following lemma that the system of equations given by (29) for all $\mathbf{x},\mathbf{z}$ has a unique solution under a regularity condition ruling out "pathological cases."

Lemma 8. *If $D_{\delta_{z_1},z_1} s_1$ is a non-periodic function of $\alpha x_j+\beta z_j$ for $j \neq 1$, then $\frac{\beta}{\alpha} = \beta$ is point-identified.*

Proof. Since $D_{\delta_{z_1},z_1} s_1$ only depends on x_j and z_j via $\alpha x_j + \beta z_j$ for $j \neq 1$, fixing all other arguments, we can write $D_{\delta_{z_1},z_1} s_1(\mathbf{x},\mathbf{z}) = h(\alpha x_j + \beta z_j)$. Consider a discrete increase δ_{z_j} in z_j and note that an increase in x_j by $\delta_{x_j} \equiv \frac{\beta}{\alpha}\delta_{z_j}$ will deliver the same change in h. Suppose by contradiction that there is a $\delta'_{x_j} \neq \frac{\beta}{\alpha}\delta_{z_j}$ that delivers the same change in h for any baseline level of $\tilde{u}_j = \alpha x_j + \beta z_j$. Then, we have $h(\tilde{u}_j + \alpha\delta'_{x_j}) = h(\tilde{u}_j + \beta\delta_{z_j})\ \ \forall \tilde{u}_j$, which holds if and only if $h(\tilde{u}_j) = h(\tilde{u}_j - \alpha\delta'_{x_j} + \beta\delta_{z_j})\ \ \forall\tilde{u}_j$, i.e. if and only if h is periodic with period $-\alpha\delta'_{x_j} + \beta\delta_{z_j} \neq 0$. Thus, if h is non-periodic, δ_{x_j} is the only increase in x_j that delivers the same change in h as an increase in z_j by δ_{z_j} for all values of (x_j, z_j), implying that $\frac{\beta}{\alpha}$ is uniquely identified as $\frac{\delta_{x_j}}{\delta_{z_j}}$. □

The simulations in Appendix D confirm that this approach delivers consistent estimates of $\frac{\beta}{\alpha}$. Here we outline the steps of this procedure:

1. Estimate the choice probability function for good 1, s_1. (Here, we focus on the case where one is using the nonparametric method of section 5.1, but alternative methods, such as flexible logit, could be used as well.)

2. For every choice set $(\mathbf{x},\mathbf{z})$ in the data, evaluate $s_1(\mathbf{x},\mathbf{z}+\delta_{z_1}+\delta_{z_j}) - s_1(\mathbf{x},\mathbf{z}+\delta_{z_j})$ for given increments δ_{z_1} and δ_{z_j}. (Consistent with the notation above, we let $\mathbf{z}+\delta_{z_1}+\delta_{z_j}$ denote a vector equal to $\mathbf{z}$ except that z_1 and z_j are increased by δ_{z_1} and δ_{z_j}, respectively).

3. Evaluate $s_1(\mathbf{x}+\delta_{x_j},\mathbf{z}+\delta_{z_1}) - s_1(\mathbf{x}+\delta_{x_j},\mathbf{z})$ and repeat this by varying δ_{x_j} over a grid of points. Find the value of δ_{x_j} that makes this difference as close as possible to the difference $s_1(\mathbf{x},\mathbf{z}+\delta_{z_1}+\delta_{z_j}) - s_1(\mathbf{x},\mathbf{z}+\delta_{z_j})$ from the previous step. Call this value $\bar{\delta}_{x_j}$.

4. Repeat steps 2-3 for every choice set $(\mathbf{x},\mathbf{z})$ in the data and take the ratio of a trimmed mean of the $\bar{\delta}_{x_j}$ values to δ_{z_j}. This is an estimate of $\frac{\beta}{\alpha}$.

Finally, note that the argument above continues to hold if we instead assume that z is continuous and x is discrete. Specifically, by Lemma 1, we have that $\frac{\partial s_1}{\partial z_1}$ depends on x_j and z_j only via $\alpha x_j + \beta z_j$. Thus, assuming that $\frac{\partial s_1}{\partial z_1}$ is a non-periodic function of $\alpha x_j + \beta z_j$, the argument in Lemma 8 with the roles of x_j and z_j flipped shows identification of $\frac{\beta}{\alpha}$ when z is continuous and x is discrete.

Regardless of whether x and z are continuous or discrete, conditioning on $\tilde{z}_j$ being the same across goods implies that consumers always maximize utility. Thus, we can continue to invoke standard full-information arguments to identify α.

A.9 Outside option known for free

Here, we show that Lemma 2 can be extended to accommodate an outside option whose utility, U_0, is known for free by consumers. Because Lemma 1 still applies in this case, we can adapt the proof in Appendix A.1 as follows:

$$\begin{aligned}
s_1 = {} & P(U_1 \geq U_k \ \forall k \in \mathcal{J} \cup \{0\}) - \sum_{\mathcal{S} \subset \mathcal{J}_1, \mathcal{S} \neq \emptyset} P\left(\{U_1 \geq U_k \forall k \in \mathcal{J} \cup \{0\}\} \cap \{EU_j \geq EU_1 \text{ for at least one } j \in \mathcal{S}\}\right. \\
& \cap \{g\left(x_1, \epsilon_1, U_j\right) \leq 0 \text{ for all } j \in \mathcal{S}\} \cap \{g\left(x_1, \epsilon_1, U_j\right) \geq 0 \text{ for all } j \in \{\mathcal{J}_1 \backslash \mathcal{S}\} \cup \{0\}\}) \\
& - P\left(\{U_1 \geq U_k \ \forall k \in \mathcal{J} \cup \{0\}\} \cap \{g(x_1, \epsilon_1, U_0) \leq 0\}\right) \\
\equiv {} & P_{4,out}\left(\tilde{\mathbf{u}}\right) - \sum_{\mathcal{S} \subset \mathcal{J}_1, \mathcal{S} \neq \emptyset} P_{5,out}^{\mathcal{S}}\left(\tilde{\mathbf{u}}, \tilde{\mathbf{eu}}, x_1\right) - P_{6,out}(\tilde{\mathbf{u}}, x_1).
\end{aligned}$$

The terms $P_{5,out}^{\mathcal{S}}$ correspond to the case where consumers fail to search good 1 because of a (relatively) high-utility inside product, whereas $P_{6,out}$ corresponds to the case where they fail to search good 1 due to the outside option. Next, for every $\mathcal{S} \subset \mathcal{J}_1, \mathcal{S} \neq \emptyset$, we have

$$\begin{aligned}
P_{5,out}^{\mathcal{S}} = {} & P\left(\{U_1 \geq U_k \ \forall k \in \mathcal{J} \cup \{0\}\} \cap \{g\left(x_1, \epsilon_1, U_j\right) \leq 0 \text{ for all } j \in \mathcal{S}\} \cap \{g\left(x_1, \epsilon_1, U_j\right) \geq 0 \text{ for all } j \in \{\mathcal{J}_1 \backslash \mathcal{S}\} \cup \{0\}\}\right) - \\
& P\left(\{U_1 \geq U_k \ \forall k\} \cap \{EU_1 \geq EU_j \text{ for all } j \in \mathcal{S}\}\right. \\
& \left. \cap \{g\left(x_1, \epsilon_1, U_j\right) \leq 0 \text{ for all } j \in \mathcal{S}\} \cap \{g\left(x_1, \epsilon_1, U_j\right) \geq 0 \text{ for all } j \in \mathcal{J}_1 \backslash \mathcal{S}\}\right) \\
= {} & P\left(\{U_1 \geq U_k \ \forall k \in \mathcal{J} \cup \{0\}\} \cap \{g\left(x_1, \epsilon_1, U_j\right) \leq 0 \text{ for all } j \in \mathcal{S}\} \cap \{g\left(x_1, \epsilon_1, U_j\right) \geq 0 \text{ for all } j \in \{\mathcal{J}_1 \backslash \mathcal{S}\} \cup \{0\}\}\right) - \\
& P\left(\{EU_1 \geq EU_j \text{ for all } j \in \mathcal{S}\} \cap \{g\left(x_1, \epsilon_1, U_j\right) \leq 0 \text{ for all } j \in \mathcal{S}\} \cap \{g\left(x_1, \epsilon_1, U_j\right) \geq 0 \text{ for all } j \in \{\mathcal{J}_1 \backslash \mathcal{S}\} \cup \{0\}\}\right) \\
\equiv {} & P_{5,1,out}^{\mathcal{S}}\left(\tilde{\mathbf{u}}, x_1\right) - P_{5,2,out}^{\mathcal{S}}\left(\tilde{\mathbf{u}}_{-1}, \tilde{\mathbf{eu}}, x_1\right),
\end{aligned}$$

where the derivation follows from the same argument as in Appendix A.1. The fact that $P_{4,out}$, $P_{5,1,out}^{\mathcal{S}}$ and $P_{6,out}$ only depend on x_j and z_j via $\tilde{u}_j$, and that $P_{5,2,out}^{\mathcal{S}}$ does not depend on z_1 implies that we can identify β using the same ratio of second derivatives as in Lemma 2.

To recover the distribution of ϵ, we need to isolate consumers who choose the utility-maximizing good. As in the baseline model, we achieve this by conditioning on choice sets where $\tilde{z}_j = \tilde{z}$ for all j. Under the assumption that consumers always search at least one of the inside goods, our model implies that consumers always search the good with the highest expected utility, which is also the good with the highest realized utility when $\tilde{z}_j$ is the same for all j. Thus, full-information arguments can be invoked to trace out the distribution of ϵ as in the baseline case.

A.10 Identification of the distribution of random coefficients

We consider the case where the coefficient β is heterogeneous across consumers with distribution F_β and study the identification of F_β. As in the homogeneous coefficients case, our argument will rely on being able to identify the good with the most desirable value of the hidden attribute. Thus, we need to assume that β_i has the same sign for all consumers (and assume it is positive without loss). Further, we assume that at least one x attribute has a homogeneous coefficient α, but allow other x attributes to have heterogeneous coefficients.[36] We start by making the following high-level assumption and then show that this holds across a range of search protocols.

Assumption 5. Let ∂_+ and ∂_- denote right and left derivatives, respectively. For some $j \neq 1$ and all $h \geq 1$,

$$\begin{aligned}
\frac{\partial^{h+1} s_1}{\partial_+ z_1 \partial_- z_j^h}(\mathbf{x}, \mathbf{z}) &= \int f_h(\mathbf{x}, \mathbf{z}; \beta)\beta^{h+1} dF_\beta(\beta) \\
\frac{\partial^{h+1} s_1}{\partial_+ z_1 \partial x_j^h}(\mathbf{x}, \mathbf{z}) &= \alpha^h \int f_h(\mathbf{x}, \mathbf{z}; \beta)\beta dF_\beta(\beta),
\end{aligned}$$

where $f_h(\mathbf{x}^*, \mathbf{z}^*; \beta)$ is constant in β and nonzero at some $(\mathbf{x}^*, \mathbf{z}^*)$.

A version of the property in Assumption 5 has been used to identify the distribution of random coefficients in full-information models (e.g., Fox, Kim, Ryan, and Bajari (2012), Allen and Rehbeck (2020)). Here, we need to ensure that good 1 continues to be the good with the highest value of the hidden attribute as we differentiate, which is why we need to take right and left derivatives with respect to z_1 and z_j, respectively. Under Assumption 5, one can immediately recover all ratios of non-central moments of β using

$$\frac{\frac{\partial^{h+1} s_1}{\partial_+ z_1 \partial_- z_j^h}(\mathbf{x}^*, \mathbf{z}^*)}{\frac{\partial^{h+1} s_1}{\partial_+ z_1 \partial x_j^h}(\mathbf{x}^*, \mathbf{z}^*)} = \frac{E(\beta^{h+1})}{E(\beta)}, \tag{30}$$

for $h \geq 2$, provided that the mean of β is nonzero (this again uses the normalization $\alpha = 1$). This mimics standard arguments for full-information models with some adjustments needed to accommodate consumer search (in particular, we have to focus on good 1 and take an extra derivative — with respect to z_1 — to get rid of "inconvenient" terms).

Assumption 5 is high-level. We now show that it holds in the search protocols in Examples 1-5. For simplicity, we consider the case with $J = 2$ goods.

Sequential search (Example 1): Let $f^{(h)}_{\epsilon_2-\epsilon_1}$ be the $h-$th derivative of the density $f_{\epsilon_2-\epsilon_1}$ and $\underline{z} = \inf \mathrm{supp}(Z)$. Then, Assumption 5 is satisfied with $f_h = (-1)^h[1 - F_{rv}(\tilde{z}_1)]f^{(h)}_{\epsilon_2-\epsilon_1}(\tilde{u}_1 - \tilde{u}_2)$ and

[36] Their distribution can be recovered by applying full-information arguments (e.g., Fox, Kim, Ryan, and Bajari (2012), Allen and Rehbeck (2020)) to choice sets where $\tilde{z}$ is the same for all goods.

$z_j^* = \underline{z}$ for all j, since at any such $\mathbf{z}$ (and any $\mathbf{x}^*$), $f_h = (-1)^h f^{(h)}_{\epsilon_2-\epsilon_1}(\alpha(x_1 - x_2))$ does not depend on β. The same argument holds for the directed cognition model (Example 2).

Satisficing (Example 3): Let $f^{(h)}_{\epsilon_2}$ be the $h-$th derivative of the density f_{ϵ_2}. Then we have $f_h = (-1)^h \int [1 - F_\tau(\epsilon_1 + \tilde{u}_1)] f^{(h)}_{\epsilon_2}(\epsilon_1 + \tilde{u}_1 - \tilde{u}_2) dF_{\epsilon_1}$, so that Assumption 5 is satisfied with $z_j^* = 0$ for all j (and any $\mathbf{x}^*$).

Full Information (Example 4): Since $f_h = (-1)^h f^{(h)}_{\epsilon_2-\epsilon_1}(\tilde{u}_1 - \tilde{u}_2)$, Assumption 5 is satisfied with any $z_1^* = z_2^*$ (and any $\mathbf{x}^*$).

Simultaneous search (Example 5): Since $f_h = (-1)^h \int \int_{\tilde{\tau}-\tilde{e u_1}}^{\infty} f^{(h)}_{\epsilon_2}(\epsilon_1 + \tilde{u}_1 - \tilde{u}_2) dF_{\epsilon_1} dF_{\tilde{\tau}}$, Assumption 5 is satisfied with $z_1^* = z_2^*, x_1^* = 0$ and any x_2^*.

These examples show that Assumption 5 holds across a range of search models. However, note that the values of $(\mathbf{x}^*, \mathbf{z}^*)$ satisfying the assumption vary with the search protocol. For example, setting $z_j^* = x_1^* = 0$ for all j works for satisficing, full information and simultaneous search, but not for sequential search. Thus, this approach requires the researcher to restrict attention to a subset of the search protocols that our main results for the homogeneous coefficients case applied to. An interesting avenue for future research would be to devise a data-driven way to find $(\mathbf{x}^*, \mathbf{z}^*)$, which would allow one to recover the distribution of random coefficients under weaker restrictions on the search protocol.

The argument above recovers ratios of the non-central moments of β. This requires either fixing one moment at a given value (which then leads to identification of all other moments) or making some further assumptions on the distribution of β. We consider three examples here (exponential, log-normal and discrete β) and show that in each of these cases, it is possible to fully recover the distribution of β. The order of the derivatives needed for identification, and thus how data demanding the approach is, will depend on the number of free parameters characterizing the distribution of β.

Lemma 9. *Suppose that Assumption 5 holds and that β is drawn from an exponential distribution with parameter λ. Then, the distribution of β is identified from derivatives of the choice probability function s_1 of order two.*

Proof. Since $E(\beta) = \frac{1}{\lambda}$ and $E(\beta^2) = \frac{2}{\lambda^2}$, the right-hand side of equation (30) with $h = 1$ equals $\frac{2}{\lambda}$, implying the conclusion. □

Next, in the log-normal case, since this has one additional parameter, it will require derivatives of up to the third order.

Lemma 10. *Suppose that Assumption 5 holds and that β is drawn from a log-normal distribution with parameters (μ, σ^2). Then, the distribution of β is identified from derivatives of the choice probability function s_1 of order up to three.*

Proof. Denote by R_1 and R_2 the left-hand side of (30) for $h = 1$ and $h = 2$, respectively. Then, we have $R_1 = \frac{e^{2\mu+2\sigma^2}}{e^{\mu+\frac{\sigma^2}{2}}}, R_2 = \frac{e^{3\mu+\frac{9}{2}\sigma^2}}{e^{\mu+\frac{\sigma^2}{2}}}$, implying $\mu = 4\log R_1 - \frac{3}{2}\log R_2$ and $\sigma^2 = \log R_2 - 2\log R_1$. □

Finally, we consider the case where β is distributed over a known grid $\beta_1, \ldots, \beta_K$. The goal is to identify the $K-1$ weights $\pi_1, \ldots, \pi_{K-1}$ (which then uniquely determine $\pi_K = 1 - \sum_{k=1}^{K-1} \pi_k$). This requires at least $K-1$ independent equations and thus higher-order derivatives.

Lemma 11. *Suppose that Assumption 5 holds and β is distributed over a known grid of strictly positive values $\beta_1, \ldots, \beta_K$ with weights $\pi_1, \ldots, \pi_K$. Then, the distribution of β is identified from derivatives of order up to K.*

Proof. Let $\beta^{\mathbf{h}} = (\beta_1^h, \ldots, \beta_K^h)$. We write (30) as $\frac{A_h}{B_h} = \frac{\beta^{\mathbf{h+1}} \cdot \pi}{\beta^{\mathbf{1}} \cdot \pi}$ and denote $\mathbf{d_h} \equiv A_h \beta^{\mathbf{1}} - B_h \beta^{\mathbf{h+1}}$, so that

$$\mathbf{d_h} \cdot \pi = 0. \tag{31}$$

Because π is a vector of probability weights, denoting $\mathbf{e} = (1, \ldots, 1)$, we have $\mathbf{e} \cdot \pi = 1$. Stacking this equation and equation (31) for $h = 1, 2, ..., K-1$ together, we obtain

$$\mathbf{D} \cdot \pi = \mathbf{e_1}, \tag{32}$$

where $\mathbf{D} \equiv [\mathbf{e}', \mathbf{d_1}', \ldots, \mathbf{d_{K-1}}']'$ and $\mathbf{e_1} = (1, 0, \ldots, 0)'$. This implies that π is identified if the matrix $\mathbf{D}$ is invertible. To this end, for any given $K \geq 2$, it can be shown that

$$det(\mathbf{D}) = (-1)^{\frac{(K-1)(K-2)}{2}} (\Pi_{h=1}^{K-1} f_h)(\Pi_{i=1}^{K} \beta_i)[\Pi_{1 \leq i < j \leq K}(\beta_i - \beta_j)] \left(\sum_{i=1}^{K} \pi_i \beta_i \right)^{K-2}.$$

using the fact that $A_h = f_h \beta^{\mathbf{h+1}} \cdot \pi, B_h = f_h \beta^{\mathbf{1}} \cdot \pi$. Therefore, the matrix $\mathbf{D}$ is invertible under the maintained assumptions. □

A.11 Utility nonlinear in x

Our practitioner's guide in Table 1 focuses on the case where the product attributes enter utility linearly. Here, we show how to extend the approach to the case where x enters utility via polynomials of degree $K_x > 1$, i.e. $u_{ij} = \sum_{k=1}^{K_x} \alpha_k x_j^k + \beta z_j + \epsilon_{ij}$. First, note that by our identification arguments:

$$\frac{\partial^2 s_1}{\partial z_1 \partial z_j}(\mathbf{x}, \mathbf{z}) \Big/ \frac{\partial^2 s_1}{\partial z_1 \partial x_j}(\mathbf{x}, \mathbf{z}) = \frac{\beta}{\sum_{k=1}^{K_x} k \alpha_k x_j^{k-1}}. \tag{33}$$

Thus, the ratio on the right-hand side can be recovered based on any sufficiently flexible estimate of s_1. Note that, unlike the case where utility is linear in x, here the denominator involves x_j. This requires changes in our estimators. For the nonparametric approach, step 2 (a) iii of Table 1 must be modified as follows: we do *not* average across all values of $(\mathbf{x}, \mathbf{z})$ in the data, but rather across values of $(\mathbf{x}, \mathbf{z})$ that all have the same x_j. For the flexible logit approach, recall that the weights w in equation (15) were chosen so that the ratio of second derivatives is a function of the coefficients and not of the specific values of $(\mathbf{x}, \mathbf{z})$. When utility is nonlinear in x, we need to relax this restriction; this could be achieved by choosing the weights in a different way, or alternatively, by including higher powers of x_k for $k \neq 1$ in equation (15). This shows that there is a trade-off: being more flexible in how x enters utility is more demanding of the data. Next, we can condition on the variance of $\tilde{z}_j$ being below a cut-off to recover the $(\alpha_k)_{k=1}^{K_x}$ coefficients using standard full-information methods (this corresponds to step 2 (c) (i) in Table 1). Finally, plugging in the estimates of $(\alpha_k)_{k=1}^{K_x}$ into (33), one can recover β.

Supplemental Appendices

Appendix B: Alternative Approaches and Support Assumptions

So far we have not focused on the support assumptions required for identification. These are nonetheless essential to understand our contribution. Alternative approaches to identification exist which differ principally in requiring much stronger support assumptions.

For instance, one could assume that the data exhibits "at-infinity" variation to effectively go back to a setting that is analogous to full information. As the expected utility for a subset of goods grows to infinity, the probability of searching those goods goes to one under reasonable assumptions on the search process. Using this, one could identify preferences using conventional arguments. However, in practice, it is often implausible that any goods are searched with probability close to 1, so this strategy would require substantial parametric extrapolation.

In contrast, our proof requires much more plausible support assumptions. There is always a good which maximizes $\tilde{z}_j$ (or our weighted index in the vector-valued case, see Appendix A.3). To recover the preference parameters in the homogeneous linear case, we only need sufficient variation to estimate second derivatives of s_1 at a single point. Our arguments do require focusing on the portion of the data where $\tilde{z}_j$ is the same across all goods. While this kind of "thin support" assumptions can lead to slow rates of convergence and thus be data-demanding, we are reassured by the fact that our approach succeeds in our empirical application of Section 7 with less than 3,000 observations. As one would expect, recovering a random coefficients distribution requires considerably more variation and data, since it involves estimating higher order derivatives of choice probabilities. We further discuss these challenges in Section 5.

Appendix C: Testing for full information with heterogeneous preferences

In Lemma 4, we considered the problem of testing the null hypothesis of full information and showed that, in the case where the coefficients α and β are homogeneous across consumers, one can test the null by checking whethe the ratios of first derivatives are attenuated relative to the ratios of second derivatives in (9). Here, we provide conditions under which the same testing approach is valid in the case where one of the two coefficients is allowed to be heterogeneous.[37] We focus on the case where β is heterogeneous and positive, i.e. all consumers agree that z is a good attribute; the arguments for the case where β is negative and the case with α is heterogeneous are analogous.[38] We also assume that the ϵ_{ij} shocks are type-I extreme-value distributed and let $s_j(\tilde{\beta})$ be the choice probability of good

[37]The reason why we let only one of the coefficients be heterogeneous is that we leverage a result from the statistics literature that applies to ratios of one-dimensional integrals.

[38]For notational simplicity, we also focus on the case where z is a scalar. The case where z is a vector follows immediately provided that one can identify good 1 appropriately.

j for consumers with $\beta = \tilde{\beta}$ under full information, i.e. $s_j(\mathbf{x}, \mathbf{z}; \tilde{\beta}) \equiv \frac{\exp(\alpha x_j + \tilde{\beta} z_j)}{\sum_{k=1}^{J} \exp(\alpha x_k + \tilde{\beta} z_k)}$.

We consider the case where the testing approach compares the ratio of second derivatives taken with respect to goods 1 and 2 to the ratio of first derivatives taken with respect to good 1. Analogous sufficient conditions could be obtained for different choices of goods. Then, we want to show that

$$\frac{\int s_1(\mathbf{x}, \mathbf{z}; \beta)(1 - s_1(\mathbf{x}, \mathbf{z}; \beta))\beta dF_\beta}{\alpha \int s_1(\mathbf{x}, \mathbf{z}; \beta)(1 - s_1(\mathbf{x}, \mathbf{z}; \beta)) dF_\beta} \geq \frac{-\int s_1(\mathbf{x}, \mathbf{z}; \beta) s_2(\mathbf{x}, \mathbf{z}; \beta)(1 - 2s_1(\mathbf{x}, \mathbf{z}; \beta))\beta^2 dF_\beta}{-\alpha \int s_1(\mathbf{x}, \mathbf{z}; \beta) s_2(\mathbf{x}, \mathbf{z}; \beta)(1 - 2s_1(\mathbf{x}, \mathbf{z}; \beta))\beta dF_\beta} \tag{34}$$

where F_β denotes the distribution of β. We take a pair $(\mathbf{x}, \mathbf{z})$ such that $\frac{\partial s_1(\mathbf{x},\mathbf{z})}{\partial x_1} > 0$ and $\frac{\partial^2 s_1(\mathbf{x},\mathbf{z})}{\partial z_1 \partial x_2} > 0$ (both of which can be verified from the data), so that (34) holds if and only if

$$\begin{aligned} - \quad & \alpha \int s_1(\mathbf{x}, \mathbf{z}; \beta) s_2(\mathbf{x}, \mathbf{z}; \beta)(1 - 2s_1(\mathbf{x}, \mathbf{z}; \beta))\beta dF_\beta \cdot \int s_1(\mathbf{x}, \mathbf{z}; \beta)(1 - s_1(\mathbf{x}, \mathbf{z}; \beta))\beta dF_\beta \geq \\ - \quad & \int s_1(\mathbf{x}, \mathbf{z}; \beta) s_2(\mathbf{x}, \mathbf{z}; \beta)(1 - 2s_1(\mathbf{x}, \mathbf{z}; \beta))\beta^2 dF_\beta \cdot \alpha \int s_1(\mathbf{x}, \mathbf{z}; \beta)(1 - s_1(\mathbf{x}, \mathbf{z}; \beta)) dF_\beta \end{aligned}$$

Then, by Theorem 2 of Wijsman (1985), the desired inequality holds if (i) $\beta > 0$, and (ii) $\frac{\alpha}{\beta}$ and $\frac{-s_1(\mathbf{x},\mathbf{z};\beta)s_2(\mathbf{x},\mathbf{z};\beta)(1-2s_1(\mathbf{x},\mathbf{z};\beta))\beta}{s_1(\mathbf{x},\mathbf{z};\beta)(1-s_1(\mathbf{x},\mathbf{z};\beta))} = -\frac{s_2(\mathbf{x},\mathbf{z};\beta)(1-2s_1(\mathbf{x},\mathbf{z};\beta))\beta}{1-s_1(\mathbf{x},\mathbf{z};\beta)}$ are monotonic functions of β in the same direction. Since we assumed throughout that $\alpha > 0$, we want to show that $-\frac{s_2(\mathbf{x},\mathbf{z};\beta)(1-2s_1(\mathbf{x},\mathbf{z};\beta))\beta}{1-s_1(\mathbf{x},\mathbf{z};\beta)}$ decreases in β monotonically. After some algebra, we have that

$$\frac{\partial \left[-\frac{s_2(\mathbf{x},\mathbf{z};\beta)(1-2s_1(\mathbf{x},\mathbf{z};\beta))\beta}{1-s_1(\mathbf{x},\mathbf{z};\beta)} \right]}{\partial \beta} < 0 \qquad \forall \beta$$

if and only if, for all β,

$$\begin{aligned} & (1 - 2s_1(\mathbf{x}, \mathbf{z}; \beta))(1 - s_1(\mathbf{x}, \mathbf{z}; \beta)) > \\ \beta \Bigg[s_1(\mathbf{x}, \mathbf{z}; \beta) \left(z_1 - \sum_{k=1}^{J} s_k(\mathbf{x}, \mathbf{z}; \beta) z_k \right) & - (1 - s_1(\mathbf{x}, \mathbf{z}; \beta))(1 - 2s_1(\mathbf{x}, \mathbf{z}; \beta)) \left(z_2 - \sum_{k=1}^{J} s_k(\mathbf{x}, \mathbf{z}; \beta) z_k \right) \Bigg] \end{aligned} \tag{35}$$

Under these conditions, at the chosen values of $\mathbf{x}, \mathbf{z}$, one can define a test that rejects the null of full information when the ratio of first derivatives is sufficiently attenuated relative to the ratio of first derivatives. Note that the condition in (35) can be verified given the support of the distribution of β. For example, if β takes values on a finite grid of points, then one needs to check whether (35) holds for all values in the grid. We emphasize that (35) and $\beta > 0$ are sufficient, but in general not necessary conditions, implying that the testing approach could be valid even if the restrictions are not satisfied.

We conclude with an example where the assumption $\beta > 0$ is violated in a way that invalidates our testing strategy.[39] Later, we will show that an alternative approach allows us to distinguish between full information with heterogeneous β and search over z with homogeneous β in this example. Suppose that there are two goods, consumers have full information, and β is equal to $\overline{\beta}$ for a fraction γ of consumers

[39] This was suggested by an anonymous referee.

and equal to zero for the remaining fraction $(1-\gamma)$. Let $s_j(\overline{\beta})$ and $s_j(0)$ be the choice probability of good j for each type of consumer, respectively, so that the overall choice probability is $s_j = \gamma s_j(\overline{\beta}) + (1-\gamma)s_j(0)$ (where we omit the dependence of the choice probabilities on the product attributes). Then, it can be shown that $\frac{\frac{\partial s_1}{\partial z_1}}{\frac{\partial s_1}{\partial x_1}} \leq \frac{\overline{\beta}}{\alpha} = \frac{\frac{\partial^2 s_1}{\partial z_1 \partial z_2}}{\frac{\partial^2 s_1}{\partial z_1 x_2}}$. In other words, due to this specific form of heterogeneity in β, the ratio of first derivatives is attenuated relative to that of second derivatives even if consumers have full information, which invalidates our testing strategy. Next, we show that an alternative approach is applicable in this case. First, note that under this null hypothesis, $\frac{\frac{\partial^2 s_1}{\partial z_1 \partial z_2}}{\frac{\partial^2 s_1}{\partial z_1 \partial x_2}} = \frac{\frac{\partial^2 s_2}{\partial z_2 \partial z_1}}{\frac{\partial^2 s_2}{\partial z_2 \partial x_1}} = \frac{\beta}{\alpha}$. On the other hand, under the alternative hypothesis that consumers search over z and have homogeneous β, these two ratios will in general be different. Specifically, (i) our main result applies to the first ratio: $\frac{\frac{\partial^2 s_1}{\partial z_1 \partial z_2}}{\frac{\partial^2 s_1}{\partial z_1 \partial x_2}} = \frac{\beta}{\alpha}$, and (ii) for the second ratio, we have $s_2 = P(U_2 > U_1) + P(U_1 > U_2, EU_2 > EU_1, g(x_1, \epsilon_1, U_2) \leq 0) \equiv P_1(\tilde{u}_1, \tilde{u}_2) + P_2(\tilde{u}_1, \tilde{u}_2, \tilde{eu}_1, \tilde{eu}_2, x_1)$, so that $\frac{\frac{\partial^2 s_2}{\partial z_2 \partial z_1}}{\frac{\partial^2 s_2}{\partial z_2 \partial x_1}} = \frac{\beta^2 \frac{\partial^2 (P_1+P_2)}{\partial \tilde{u}_2 \partial \tilde{u}_1}}{\alpha\beta\left(\frac{\partial^2 (P_1+P_2)}{\partial \tilde{u}_2 \partial \tilde{u}_1} + \frac{\partial^2 P_2}{\partial \tilde{u}_2 \partial \tilde{eu}_1}\right) + \beta \frac{\partial^2 P_2}{\partial \tilde{u}_2 \partial x_1}}$. Because $\frac{\partial^2 P_2}{\partial \tilde{u}_2 \partial \tilde{eu}_1}$ and $\beta \frac{\partial^2 P_2}{\partial \tilde{u}_2 \partial x_1}$ are generally nonzero, the expression above is in general not equal to $\frac{\beta}{\alpha}$. This example illustrates that it is possible to distinguish empirically between the hypothesis that consumers are fully informed with some of them having a zero coefficient on z and the competing hypothesis that they search over z and have homogeneous preferences over that attribute.

Appendix D: Simulations Results

To test the performance of our approach, we consider several simulations. In all simulations, we generate $N = 20,000$ choices with utility given by:

$$U_{ij} = \alpha x_{ij} + \beta z_{ij} + \epsilon_{ij} \tag{36}$$

with $\alpha = \beta = 1$, $x_{ij} \sim_{i.i.d} N(0,1)$, and ϵ_{ij} i.i.d. Type 1 extreme value. We separately consider the case where z_{ij} is continuous ($z_{ij} \sim_{i.i.d.} N(0,1)$) and the case where it is discrete (distributed uniformly over a grid of ten equally spaced points from 0.1 to 0.9). Further, we set $\gamma_1 = 0$, i.e. we let consumers correctly infer that x_j is not informative about z_j.

We simulate data from four data generating processes, three of which satisfy the assumptions of our lemma and one of which does not. These are:

1. Weitzman search, with search costs $c \sim LogNormal(-2, 2.25)$;
2. Satisficing, i.e. searching in order of expected utility until utility-in-hand is at least $T \sim LogNormal(-0.35, 2.25)$;
3. Search all goods with expected utility above a threshold given by $c \sim N(-1, 16)$ (if no goods are above the threshold, search and choose the good with the highest expected utility);

4. Randomly search $K \in \{1, \dots, J\}$ goods, where the searched goods are the K highest in terms of expected utility.

DGPs 1-3 satisfy our assumptions. By contrast, DGP 4 violates Assumption 2(ii) because the decision of whether to search a good does not just depend on that good's expected utility, but on the expected utilities of all goods.

Bernstein Polynomial Simulation Results Table 7 reports results from the Bernstein approximation of the second-derivative ratio which identifies β/α. For comparison, we also report estimates of $\frac{\partial s_j/\partial z_j}{\partial s_j/\partial x_j}$, which would recover β/α with full information. In all cases, the estimates based on first-derivatives are attenuated relative to the true values. This occurs for the reason discussed in Section 2: consumer insensitivity to variation in z for goods that are not searched biases the coefficients towards zero. In contrast, the confidence intervals from Bernstein estimation of the second-derivative ratio include the true values in DGPs 1-3, and are fairly precise for the three-good case. For DGP 4, where the assumptions of our model do not hold (see Appendix A.7), the coefficient is attenuated in the three-good case, although the point estimates remain much closer to the true values relative to the first-derivative estimates.

Table 7: Bernstein Approximation

	Two goods				Three goods	
	Continuous z		Discrete z		Continuous z	
DGP	First Deriv	Second Deriv	First Deriv	Second Deriv	First Deriv	Second Deriv
1	0.61	0.98	0.34	0.77	0.40	1.00
	(0.02)	(0.30)	(0.10)	(0.31)	(0.01)	(0.08)
2	0.69	1.28	0.55	1.04	0.36	0.94
	(0.02)	(0.54)	(0.11)	(0.27)	(0.01)	(0.07)
3	0.53	0.87	0.84	1.15	0.33	0.87
	(0.02)	(0.19)	(0.12)	(0.25)	(0.01)	(0.07)
4	0.44	0.80	0.58	1.04	0.21	0.63
	(0.02)	(0.30)	(0.08)	(0.22)	(0.01)	(0.08)

Note: Across all rows, the sample size is $N = 20,000$ and the data in each row is generated by the corresponding DGP described in the main text. In all cases, the true value is 1. Standard deviations across 250 simulations are reported in parentheses. "First Deriv" refers to the standard approach based on first derivatives and "Second Deriv" refers to our approach based on second derivatives.

Flexible Logit Simulation Results For each of the DGPs described above, we consider simulations with $J \in \{2, 3, 5, 10\}$. We report estimates from the flexible logit model as well as the standard logit model. We bootstrap the standard errors using 250 repetitions.

Results from these simulations are reported in Table 8. The table shows estimates of β/α from a conditional logit model with no adjustment for imperfect information, as well as the second-derivative ratio estimates from the flexible logit model. In the standard logit model, the coefficient is attenuated, typically biased towards zero by 30-50%. The flexible logit model performs substantially better, with 95% confidence intervals including the true estimates in DGPs 1-3. Perhaps surprisingly, the flexible logit model also performs well for DGP 4; the confidence intervals include the true values for 2, 3 and 10 goods (and almost for 5 goods), and have less bias than the standard logit model regardless of the number of goods.

Table 8: Estimator Based on Second-derivatives Ratio (Flexible Logit) vs Standard Logit

	Number of Goods							
	2		3		5		10	
DGP	Standard	Flexible	Standard	Flexible	Standard	Flexible	Standard	Flexible
1	0.6590	1.0162	0.6330	1.0809	0.6050	1.0872	0.5770	1.0249
	(0.0158)	(0.1314)	(0.0122)	(0.1239)	(0.0095)	(0.1508)	(0.0089)	(0.1481)
2	0.7403	1.0251	0.6194	1.0277	0.4587	1.0508	0.2909	1.1817
	(0.0162)	(0.1151)	(0.0135)	(0.1139)	(0.0102)	(0.1571)	(0.0083)	(0.4269)
3	0.5424	1.0522	0.5945	1.0397	0.6543	0.9579	0.7246	0.8987
	(0.0149)	(0.1725)	(0.0117)	(0.1408)	(0.0099)	(0.1250)	(0.0106)	(0.1038)
4	0.4543	1.0720	0.5568	0.9644	0.6691	0.8609	0.7887	0.8449
	(0.0140)	(0.1850)	(0.0118)	(0.1631)	(0.0105)	(0.1179)	(0.0104)	(0.0945)

Note: Across all rows, the sample size is $N = 20,000$ and the data in each row is generated by the corresponding DGP described in the main text. "Standard" refers to estimates of β/α from a conventional logit model, and "Flexible" refers to estimates from the flexible logit model. In all cases, the true value is 1. Standard deviations across 250 simulations are reported in parentheses.

Appendix E: Derivation of Flexible Logit Weights and Choice Probabilities

To motivate our "flexible logit" approach to estimating $s_1(\mathbf{x}, \mathbf{z})$, note that standard full-information logit models typically impose strong restrictions on the structure of the derivatives of choice probabilities. Specifically, if $u_{ij} = v_j + \epsilon_{ij}$ and ϵ_{ij} is i.i.d. type-I extreme value, where v_j is a differentiable function of x_j and z_j, then for $q_j \in \{x_j, z_j\}$:

$$\begin{aligned} \frac{\partial s_j}{\partial q_j} &= \frac{\partial s_j}{\partial v_j}\frac{\partial v_j}{\partial q_j} = \frac{\partial v_j}{\partial q_j} s_j(1-s_j) \\ \frac{\partial s_j}{\partial q_{j'}} &= \frac{\partial s_j}{\partial v_{j'}}\frac{\partial v_{j'}}{\partial q_{j'}} = -\frac{\partial v_{j'}}{\partial q_{j'}} s_j s_{j'} \\ \frac{\partial^2 s_j}{\partial z_j \partial q_{j'}} &= -\frac{\partial v_{j'}}{\partial q_{j'}}\frac{\partial v_j}{\partial z_j} s_j s_{j'}(1-2s_j) \end{aligned} \tag{37}$$

for $j' \neq j$. Thus, in a conventional logit model, $\frac{\partial^2 s_1}{\partial z_1 \partial z_{j'}} / \frac{\partial^2 s_1}{\partial z_1 \partial x_{j'}} = \frac{\partial s_1}{\partial z_{j'}} / \frac{\partial s_1}{\partial x_{j'}} = \frac{\partial v_{j'}}{\partial z_{j'}} / \frac{\partial v_{j'}}{\partial x_{j'}}$ for all $j' \neq 1$, and this further equals $\frac{\partial s_1}{\partial z_1} / \frac{\partial s_1}{\partial x_1}$ when $\frac{\partial v_j}{\partial q_j} = \frac{\partial v_{j'}}{\partial q_{j'}}$ for all j, j'. We would like to estimate a model of s_1 which is sufficiently flexible that ratios of first-derivatives differ from ratios of second derivatives, as will generally occur if consumers engage in search. To allow for this additional flexibility, we let the utility for good 1 depend directly on attributes of rival goods as follows:

$$v_1 = ax_1 + b_1 z_1 + \sum_{k \neq 1} \left(\eta_k w_{z1k} z_k + \eta_{2k} w_{x1k} x_k + \rho_k w_{z2k} z_k z_1 + \rho_{2k} w_{x2k} x_k z_1 \right) \tag{38}$$

where w_{z1k}, w_{x1k}, w_{z2k} and w_{x2k} are known weights, and a, b_1, η_k, η_{2k}, ρ_k and ρ_{2k} are coefficients to be estimated. Further, we let $v_k = ax_k + bz_k$ for $k \neq 1$.

In this section, we derive the relevant derivatives of choice probabilities for the flexible logit model described in the text and motivate our choice of weights. The weights w_{x1k}, w_{z1k}, w_{x2k} and w_{z2k} are chosen so that, given the logit functional form, $\frac{\partial^2 s_1}{\partial z_1 \partial z_j} / \frac{\partial^2 s_1}{\partial z_1 \partial x_j}$ can be constant across goods as our structural model implies when these weights are regarded as constant in derivatives. With these weights, we have the following derivatives :

$$
\begin{aligned}
\frac{\partial v_1}{\partial z_1} &= b_1 + \sum_{k \neq 1} (\rho_k w_{z2k} z_k + \rho_{2k} w_{x2k} x_k) \\
\frac{\partial s_1}{\partial x_1} &= \frac{\partial s_1}{\partial v_1} \frac{\partial v_1}{\partial x_1} = a s_1 (1 - s_1) \\
\frac{\partial s_1}{\partial z_1} &= \frac{\partial s_1}{\partial v_1} \frac{\partial v_1}{\partial z_1} = \frac{\partial v_1}{\partial z_1} s_1 (1 - s_1) \\
\frac{\partial s_1}{\partial x_{j'}} &= \frac{\partial s_1}{\partial v_{j'}} \frac{\partial v_{j'}}{\partial x_{j'}} + \frac{\partial s_1}{\partial v_1} \frac{\partial v_1}{\partial x_{j'}} = -a s_1 s_{j'} + [\eta_{2j'} w_{x1j'} + \rho_{2j'} w_{x2j'} z_1] s_1 (1 - s_1) \\
\frac{\partial s_1}{\partial z_{j'}} &= \frac{\partial s_1}{\partial v_{j'}} \frac{\partial v_{j'}}{\partial z_{j'}} + \frac{\partial s_1}{\partial v_1} \frac{\partial v_1}{\partial z_{j'}} = -b s_1 s_{j'} + [\eta_{j'} w_{z1j'} + \rho_{j'} w_{z2j'} z_1] s_1 (1 - s_1)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 s_1}{\partial z_1 \partial x_{j'}} &= \frac{\partial^2 s_1}{\partial v_1 \partial x_{j'}} \frac{\partial v_1}{\partial z_1} + \frac{\partial s_1}{\partial v_1} \frac{\partial^2 v_1}{\partial z_1 \partial x_{j'}} \\
&= \frac{\partial v_1}{\partial z_1} (1 - 2 s_1) \frac{\partial s_1}{\partial x_{j'}} + s_1 (1 - s_1) \rho_{2j'} w_{x2j'} \\
\frac{\partial^2 s_1}{\partial z_1 \partial z_{j'}} &= \frac{\partial^2 s_1}{\partial v_1 \partial z_{j'}} \frac{\partial v_1}{\partial z_1} + \frac{\partial s_1}{\partial v_1} \frac{\partial^2 v_1}{\partial z_1 \partial z_{j'}} \\
&= \frac{\partial v_1}{\partial z_1} (1 - 2 s_1) \frac{\partial s_1}{\partial z_{j'}} + s_1 (1 - s_1) \rho_{j'} w_{z2j'}
\end{aligned} \tag{39}
$$

And also:

$$
\begin{aligned}
\frac{\partial^2 s_1}{\partial z_1 \partial z_{j'}} / \frac{\partial^2 s_1}{\partial z_1 \partial x_{j'}} &= \frac{\frac{\partial v_1}{\partial z_1} (1 - 2 s_1) \frac{\partial s_1}{\partial z_{j'}} + s_1 (1 - s_1) \rho_{j'} w_{z2j'}}{\frac{\partial v_1}{\partial z_1} (1 - 2 s_1) \frac{\partial s_1}{\partial x_{j'}} + s_1 (1 - s_1) \rho_{2j'} w_{x2j'}} \\
&= \frac{\frac{\partial v_1}{\partial z_1} (1 - 2 s_1) \left(-b s_1 s_{j'} + [\eta_{j'} w_{z1j'} + \rho_{j'} w_{z2j'} z_1] s_1 (1 - s_1) \right) + s_1 (1 - s_1) \rho_{j'} w_{z2j'}}{\frac{\partial v_1}{\partial z_1} (1 - 2 s_1) \left(-a s_1 s_{j'} + [\eta_{2j'} w_{x1j'} + \rho_{2j'} w_{x2j'} z_1] s_1 (1 - s_1) \right) + s_1 (1 - s_1) \rho_{2j'} w_{x2j'}}
\end{aligned} \tag{40}
$$

If we define the weights: $w_{x1j'} = w_{z1j'} = \frac{s_{j'}}{1 - s_1}$ and $w_{x2j'} = w_{z2j'} = [\frac{z_1(1-s_1)}{s_{j'}} + \frac{(1-s_1)}{(\partial v_1/\partial z_1)(1-2s_1)s_{j'}}]^{-1} = \frac{(1-2s_1)s_{j'}}{1-s_1} \left(\frac{1}{\partial v_1/\partial z_1} + (1 - 2s_1) z_1 \right)^{-1} = \frac{(\partial v_1/\partial z_1)(1-2s_1)s_{j'}}{1-s_1} (1 + (1 - 2s_1) z_1 (\partial v_1/\partial z_1))^{-1}$, then:

$$
\begin{aligned}
\frac{\partial^2 s_1}{\partial z_1 \partial z_{j'}} / \frac{\partial^2 s_1}{\partial z_1 \partial x_{j'}} &= \frac{\frac{\partial v_1}{\partial z_1} (1 - 2 s_1) s_1 s_{j'} \left(-b + \eta_{j'} w_{z1j'} \frac{(1-s_1)}{s_{j'}} + \rho_{j'} w_{z2j'} [\frac{z_1(1-s_1)}{s_{j'}} + \frac{(1-s_1)}{(\partial v_1/\partial z_1)(1-2s_1)s_{j'}}] \right)}{\frac{\partial v_1}{\partial z_1} (1 - 2 s_1) s_1 s_{j'} \left(-a + \eta_{j'} w_{x1j'} \frac{(1-s_1)}{s_{j'}} + \rho_{j'} w_{x2j'} [\frac{z_1(1-s_1)}{s_{j'}} + \frac{(1-s_1)}{(\partial v_1/\partial z_1)(1-2s_1)s_{j'}}] \right)} \\
&= \frac{\frac{\partial v_1}{\partial z_1} (1 - 2 s_1) s_1 s_{j'} \left(-b + \eta_{j'} + \rho_{j'} \right)}{\frac{\partial v_1}{\partial z_1} (1 - 2 s_1) s_1 s_{j'} \left(-a + \eta_{2j'} + \rho_{2j'} \right)} \\
&= \frac{-b + \eta_{j'} + \rho_{j'}}{-a + \eta_{2j'} + \rho_{2j'}}
\end{aligned} \tag{41}
$$

Thus, we have:

$$\frac{\partial^2 s_1}{\partial z_1 \partial z_{j'}} \bigg/ \frac{\partial^2 s_1}{\partial z_1 \partial x_{j'}} \quad = \quad \frac{-b + \eta_{j'} + \rho_{j'}}{-a + \eta_{2j'} + \rho_{2j'}} \tag{42}$$

where $w_{z1j'} = w_{x1j'} = \frac{s_{j'}}{1-s_1}$ and $w_{x2j'} = w_{z2j'} = (\partial v_1/\partial z_1)\frac{(1-2s_1)s_{j'}}{1-s_1}\left(1 + (\partial v_1/\partial z_1)(1-2s_1)z_1\right)^{-1}$. This implies that the above ratio is a constant for each j'.

Estimation of the model with these weights is infeasible since the levels of the choice probabilities s_1 and s_k, as well as the derivatives $\partial v_1/\partial z_1$ are unknown ex ante and thus we do not know the weights. We estimate the model via a two-step process where s_1 and s_k are estimated using a standard logit model (where utility for each good is a linear function of x_j and z_j), these estimates are used to construct weights, and then the model in equation (38) is estimated treating these weights as constants.[40]

To recover estimates of β/α from the flexible logit model, we use the ratio in equation (42). In cases where the identity of goods is not meaningful (e.g. "good 2" does not refer to the same good across different choice sets and there are no alternative-specific fixed effects), we can further impose $\eta_k = \eta$, $\eta_{2k} = \eta_2$, $\rho_k = \rho$ and $\rho_{2k} = \eta_2$, which gives a single estimate of $\frac{\beta}{\alpha}$.

Appendix F: Recovery of Search Costs Given Preferences in the Weitzman Model

Suppose that consumers search sequentially according to the model of Weitzman (1979). For simplicity, here we consider the case where $\gamma_1 = 0$.

As shown in Armstrong (2017),[41] the optimal search strategy is for consumers to behave as if they were choosing among options in a static model with utilities given by $\tilde{U}_{ij} = x_j\alpha + \min\{z_j, rv_i\}\beta + \epsilon_{ij}$, where rv_i denotes i's reservation value in units of z (see Example 1). Thus, dropping i subscripts, ordering goods so that $z_1 \geq z_2 \geq \ldots \geq z_J$, and letting

$$E_t \equiv \{\epsilon : \epsilon_k - \epsilon_1 \leq (x_1 - x_k)\,\alpha, k = 2, ..., J-t-1\} \cap \{\epsilon : \epsilon_h - \epsilon_1 \leq (x_1 - x_h)\,\alpha + (rv - z_h)\,\beta, h = J-t, ..., J\}$$

[40] Since $\partial v_1/\partial z_1$ is estimated imprecisely from the standard logit, when $1 + (\partial v_1/\partial z_1)(1-2s_1)z_1$ is close to 0 (leading to very large weights), we set $\partial v_1/\partial z_1 = 0$ when the former term falls below 1 in absolute value.

[41] See also Choi, Dai, and Kim (2018).

we can write

$$\begin{aligned}
s_1 &= P\left(x_1\alpha + \min\left\{z_1, rv\right\}\beta + \epsilon_1 \geq x_k\alpha + \min\left\{z_k, rv\right\}\beta + \epsilon_k \ \forall k\right) \\
&= P\left(\epsilon_k - \epsilon_1 \leq \left(x_1 - x_k\right)\alpha \ \forall k\right) \cdot P\left(rv \leq z_J\right) \\
&+ \sum_{t=0}^{J-2} \int P\left(\left\{\epsilon \in E_t\right\} \cap \left\{z_{J-t} \leq rv \leq z_{J-t-1}\right\}\right) dF_{rv}\left(rv\right) \\
&+ P\left(\epsilon_k - \epsilon_1 \leq \left(x_1 - x_k\right)\alpha + \left(z_1 - z_k\right)\beta \ \forall k\right) \cdot P\left(rv \geq z_1\right)
\end{aligned}$$

where F_{rv} denotes the cdf of rv and the second equality assumes that search costs (and thus rv) are independent of ϵ. Therefore, we have

$$\frac{\partial s_1}{\partial z_1} = \left[\frac{\partial}{\partial z_1} P\left(\epsilon_k - \epsilon_1 \leq \left(x_1 - x_k\right)\alpha + \left(z_1 - z_k\right)\beta \ \forall k\right)\right] P\left(rv \geq z_1\right) \tag{43}$$

Given identification of (α, β) by the argument in Section 2, the first term on the RHS of (43) is identified if we assume a distribution for ϵ.[42] Thus, $P\left\{rv \geq z_1\right\}$ is identified. Repeating the argument for all z_1, one can trace out the entire distribution of rv. Since c, the search cost for consumer i, is a known transformation of rv,[43] the distribution of c is also identified.

Equation (43) also lends itself to a different argument that does not require assuming a distribution for ϵ, but instead relies on "at-infinity" variation. Note that the first term on the RHS of (43) is invariant to increasing all z_j's by the same amount. Thus, we can write

$$\frac{\frac{\partial s_1}{\partial z_1}\left(\mathbf{z} + \mathbf{\Delta}\right)}{\frac{\partial s_1}{\partial z_1}\left(\mathbf{z}\right)} = \frac{P\left(rv \geq z_1 + \Delta\right)}{P\left(rv \geq z_1\right)} \tag{44}$$

where $\mathbf{\Delta}$ is a $J-$vector with all elements equal to some Δ. Letting $\Delta \to -\infty$, the numerator on the RHS of (44) goes to 1, which yields identification of $P\left(rv \geq z_1\right)$. Repeating the argument for all z_1, one can trace out the entire distribution of rv and recover the distribution of c as above.

There are a few reasons researchers might be interested in recovering search costs given the methods developed here. A full normative evaluation of an information intervention might directly include search costs: information may benefit consumers both by helping them make better choices and by helping them make choices more easily, and search costs quantify the latter effect. Note that structural search costs may be the wrong object to use for normative evaluation even if a structural search model performs well as a positive model of choices. For example, if consumers spend one hour choosing

[42] In Section 2, we could normalize $\alpha = 1$ since we were treating the distribution of ϵ nonparametrically. Here, we assume a distribution for ϵ (e.g., Gumbel) and thus we are not free to normalize α. However, note that an argument analogous to that of Section 2 immediately yields identification of α. Specifically, when $z_j = z_k$ for all j, k, consumers never make a mistake and thus variation in $\mathbf{x}$ identifies α using standard full information methods.

[43] This assumes that the prior F_z used by consumers in forming expectations are known to the researcher, as in the case where consumers have rational expectations and F_z coincides with the observed distribution of z across goods and/or markets.

insurance plans and we estimate that they act *as if* they have search costs of \$1,000 per plan, this does not imply that they are made \$1,000 better off by eliminating the need to search. In other words, search behavior may be well-described by a model with large search costs even if consumers' willingness to pay to avoid search is substantially less than the costs implied by any given model. Back of the envelope estimates of search costs based on survey data or other information on the time consumers spend choosing may be more credible and less prone to misspecification than structural estimates (e.g. Kling, Mullainathan, Shafir, Vermeulen, and Wrobel (2008)).

Search costs may also be of interest for counterfactuals where the choice environment is altered in ways that change search behavior without eliminating search entirely. As we emphasize above, eliminating search entirely is a reasonable counterfactual in our setting where search uncovers objective information that is available to the econometrician. However, other counterfactuals may be of interest, such as changing the order in which items are presented to consumers in search (Compiani, Lewis, Peng, and Wang 2024). This typically requires placing more structure on the search process.

Appendix G: Welfare Benefits of Information

We first consider the case with an arbitrary distribution of unobservables: $U_{ij} = \alpha x_j + \beta z_j + \epsilon_{ij}$ (later, we specialize things to the logit case). Here, we assume that the choices in the status quo where consumers may not be fully informed can be represented by a positive utility function of the form $U_{ij,pos} = \alpha_{pos} x_j + \beta_{pos} z_j + \epsilon_{ij}$, i.e. that consumer i chooses option j in the status quo if and only if $U_{ij,pos} \geq U_{ik,pos}$ for all $k \neq j$. The assumption that $U_{ij,pos}$ is separable in ϵ_{ij} is not without loss and can be microfounded given a model of search. For example, in the Weitzman-type model of Example 1, one can use the argument in Armstrong (2017) to show that choices are represented by a set of indices that are separable in ϵ_{ij} (under our maintained assumption that ϵ_{ij} is only revealed via search). Then, the consumer surplus in the status quo is:

$$E(CS_0) = \sum_j \int_{\tilde{M}_j} (\alpha x_j + \beta z_j + \epsilon_{ij}) dF(\epsilon_i) \tag{45}$$

where F denotes the distribution of ϵ_i and $\tilde{M}_j = \{\epsilon_i : U_{ij,pos} \geq U_{ik,pos} \ \forall k \neq j\}$. Note that $E(CS_0)$ is identified under our assumptions. Specifically, Lemma 2 shows that β as well as the distribution of ϵ_i are identified (with the normalization $\alpha = 1$). Further, given the distribution of ϵ_i, α_{pos} and β_{pos} are identified from the data (which by definition is generated under the status quo) under standard arguments for full information models (Matzkin 1993).

Next, the consumer surplus when consumers are fully informed is given by:

$$E(CS_1) = \sum_j \int_{M_j} (\alpha x_j + \beta z_j + \epsilon_{ij}) dF(\epsilon_i) \tag{46}$$

where $M_j = \{\epsilon_i : U_{ij} \geq U_{ik} \ \forall k \neq j\}$. This quantity is immediately identified given identification of normative preferences in Lemma 2 without any additional assumptions. Combining the arguments above identifies the welfare benefits of full information, $E(CS_1) - E(CS_0)$.

In practice, we make extra parametric assumptions in applications. To deal with the curse of dimensionality, we assume that the ϵ shocks are i.i.d. Gumbel. In order to approximate status quo choice probabilities, we estimate a standard logit model on the data where consumers are (possibly) uninformed (the flexible logit estimates could be used here as well to provide additional flexibility). Similarly, we approximate informed choices using this same logit model with the flexible logit estimate of β replacing the standard logit estimate. Finally, we use this logit model (with the flexible logit estimate of β) to evaluate normative choices, consistent with our assumption that consumers fully learn their utilities through search.

To spell things out in more detail, denote by β_{pos} the standard logit estimate of β and by β_{norm} the flexible logit estimate, and let $s_j(\beta_{pos})$ be product j's choice probability when consumers attach a weight of β_{pos} to the z attribute. Then, as shown in Appendix D of Abaluck and Gruber (2009), we can compute the dollar-equivalent consumer surplus of potentially uninformed consumers as:

$$E(CS_0) = -\frac{1}{\alpha_p}\left[\sum_k (\beta_{norm} z_k - \beta_{pos} z_k) s_k(\beta_{pos}) + \ln \sum_k \exp(\alpha x_k + \beta_{pos} z_k)\right]$$

where α_p is the (normative) marginal utility of income, estimated as the coefficient on price (consistent with the settings in our applications, we assume price is visible to consumers at no cost, i.e. that it is one of the x attributes). On the other hand, consumer surplus under full information is given by the conventional log-sum formula:

$$E(CS_1) = -\frac{1}{\alpha_p} \ln \sum_k \exp(\alpha x_k + \beta_{norm} z_k),$$

The change in consumer surplus from providing consumers with information is thus:

$$\Delta CS = -\frac{1}{\alpha_p}\left[\ln \sum_k \exp(\alpha x_k + \beta_{norm} z_k) - \ln \sum_k \exp(\alpha x_k + \beta_{pos} z_k) - \sum_k (\beta_{norm} z_k - \beta_{pos} z_k) s_k(\beta_{pos})\right].$$

This is the quantity we report in both of our applications.

Appendix H: Details of Estimation in the Laboratory Experiment

Our pre-registered approach is as follows: we approximate the function s_1 by taking the tensor product of univariate Bernstein polynomials, one for each argument of the function.[44] Further, we impose the natural constraint that s_1 be decreasing in the price of book 1 and the discounts of books 2 and 3, and increasing in the discount of book 1 and the prices of books 2 and 3. The main result of this procedure is an estimate of β/α_1, which we obtain by dividing a trimmed mean (across choices) of $\frac{\partial^2 s_1}{\partial discount_1 \partial discount_j}$ by a trimmed mean of $\frac{\partial^2 s_1}{\partial discount_1 \partial price_j}$ for all $j \neq 1$, and then averaging across j.[45] Note that, unlike in Lemma 2, here we will assume a distribution for ϵ (to be used in our counterfactual analysis below) and thus we are not free to impose a scale normalization on α_1. Instead, we estimate α by selecting choice sets where the across-product variance of discounts is in the bottom quartile[46] of the distribution across choice sets and estimating a standard logit model (where we drop the discount from the list of explanatory variables). Further, because prices and discounts are randomized — and participants are explicitly told so — we assume that no inference on discounts is made based on prices, i.e. we set $\gamma_1 = 0$.

Appendix I: Testing the Expected Utility Assumption in the Laboratory Experiment

As discussed in Section 4.3, while the expected utility assumption cannot be verified directly, it can be tested along with the other restrictions of our model. One such test is to compute bounds on the choice probabilities implied by the model. Given our estimates of preferences and assumptions about the distribution of ϵ_{ij}, we can compute the upper and lower bounds described in Section 4.3 for each individual via simulation. We sort the data by the lower bound, bin the data into 100 quantiles, and graph in each quantile the mean of the upper and lower bounds, as well as the choice probabilities estimated via Bernstein polynomials.

Figure 4 shows the results of this exercise. We can see that the bounds in the experimental data have some bite: the range between the lower bound and the upper bound ranges from 13 to 20 percentage points. The estimated choice probabilities in nearly all cases lie within this range. These probabilities thus appear broadly consistent with the expected utility assumption.

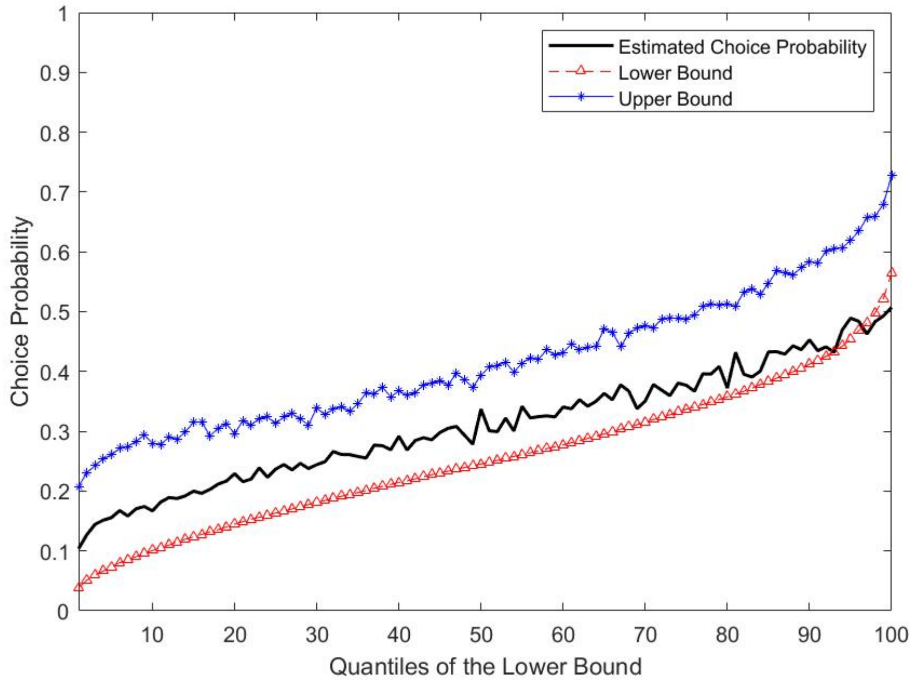
---

[44] We use univariate polynomials of degree three for the arguments z_1, x_2, z_2 and of degree two for the remaining arguments. The total degree of the approximation is 15. We chose these values for the polynomial degrees since we found that they worked well in simulations with a similar number of goods and across a range of search protocols. One could use recent results by Chen, Christensen, and Kankanala (2024) for a more formal approach.

[45] Specifically, for each second derivative, we take the mean over values in the interquartile range. As is often the case in nonparametric estimation, trimming helps obtain less noisy estimates. Again, the amount of trimming was chosen based on simulation results.

[46] Table 13 shows that the results are robust to the choice of this threshold.

Figure 4: Choice Probabilities, Upper and Lower Bounds from Expected Utility Assumption



Appendix J: Field Validation Details

J.1 Data Cleaning

The dataset from Kaggle.com contains 9,917,530 observations at the hotel-consumer level. We filtered out the following categories of observations. First, the data set contains some errors in the price information. We removed search impressions that contain at least one observation for which the listed hotel price is below \$10 or above \$1000 per night, or the implied tax paid per night either exceeds 30% of the listed hotel price, or is less than \$1. Second, we removed the search impressions where the consumer observed a hotel in position 5, 11, 17 or 23. These positions usually correspond to "opaque offers" (Ursu (2018) provides a detailed description of this feature in the data). Third, the original data set contains observations on more than 20,000 destinations, with a median of two search impressions per destination. We focused our attention on destinations with at least 50 search impressions. Fourth, we kept the search impressions where transactions happened within the top 10 positions excluding the opaque offer positions, and we only kept these top 10 hotels in these choice sets. The final dataset then contains 54,648 choice sets and 546,480 observations. Table 9 provides a detailed description of each variable.

J.2 Confidence Interval Construction

We follow these steps:

1. For every bootstrap sample (indexed by $n = 1, \ldots, 250$) and every choice of the x variable

Table 9: Variable Description

Variable	Description
Price	Gross price in USD
Stars	Number of hotel stars
Review Score	User review score, mean over sample period
Chain	Dummy for whether hotel is part of a chain
Location Score	Expedia's score for desirability of hotel's location
Promotion	Dummy for whether hotel is on promotion

(indexed by k), compute an estimate $\hat{\beta}_{k,n} = \left(\frac{\hat{\beta}}{\alpha}\right)_{k,n} \hat{\alpha}_{k,n}$ using steps 1-5 in Table 1. The possible choices of x depend on the model. For example, in Model I where the candidate z is the location score, the possible x variables are price, stars and review score.[47] Repeat the same procedure using the original sample to obtain point estimates β_k.

2. Compute the variance of $\hat{\beta}_{k,n}$ across bootstrap samples, denoted by $var(\hat{\beta}_k)$.

3. For each n, calculate the weighted average $\hat{\beta}_n = \sum_k w_k \cdot \hat{\beta}_{k,n}$, where the weights, $w_k = \frac{1/var(\hat{\beta}_k)}{\sum_k (1/var(\hat{\beta}_k))}$, are proportional to the inverse of variance so that we put less weight on less informative estimates (step 5 in Table 1). Repeat the same procedure using the original sample to obtain the point estimate $\hat{\beta}$.

4. Compute the bias-corrected confidence interval (Davison and Hinkley 1997) as follows. Let $z_0 = \Phi^{-1}\{\#(\hat{\beta}_n \leq \hat{\beta})/250\}$, where $\#(\hat{\beta}_n \leq \hat{\beta})$ is the number of elements of the bootstrap distribution that are less than or equal to the estimate from the original dataset and Φ is the standard normal cdf. Let $p_1 = \Phi\left(2z_0 - z_{1-\alpha/2}\right)$, $p_2 = \Phi\left(2z_0 + z_{1-\alpha/2}\right)$, where $z_{1-\alpha/2}$ is the $(1-\alpha/2)-$th quantile of the standard normal distribution. Compute the bias-corrected $(1-\alpha)\%$ confidence interval $[\beta^*_{p_1}, \beta^*_{p_2}]$, where β^*_p is the pth quantile of the bootstrap distribution $(\hat{\beta}_1, ..., \hat{\beta}_{250})$ from step 3.

J.3 Estimation Results

Tables 10 and 11 show detailed results from the flexible logit and standard logit estimations. To facilitate comparisons, we report the coefficients multiplied by the standard deviation of the corresponding variable.

[47] We estimate $\alpha_{k,n}$ via a logit model using only the choice sets where the variance of the z across products is in the bottom decile. The results are robust to this threshold (see Table 13).

Table 10: Estimation Results: Normalized β Estimates

z Variable	Standard Estimate	Standard CI	Flexible Estimate	Flexible CI
Location Score	0.298	(0.278, 0.317)	0.770	(0.548, 1.107)
Price	-1.085	(-1.109, -1.061)	-1.192	(-1.352, -1.044)
Review Score	0.172	(0.159, 0.185)	0.450	(0.083, 0.641)
Stars	0.386	(0.369, 0.403)	0.256	(0.088, 0.380)

Note: We report point estimates and 95% confidence intervals for the β coefficients for different choices of the z variable. The first two columns report results from the standard logit model and the second two report results from the flexible logit approach.

Table 11: Estimation Results: Difference in Magnitude of β Estimates

z Variable	Point Estimate	Confidence Interval
Location Score	0.472	(0.258, 0.783)
Price	0.107	(-0.032, 0.256)
Review Score	0.278	(-0.092, 0.486)
Stars	-0.131	(-0.271, 0.007)

Note: For different choices of the z variable, we report point estimates and 95% confidence intervals for the difference between the absolute value of the β coefficient estimate from the flexible logit approach and the absolute value of the estimate from the standard logit model.

J.4 Robustness Checks for Expedia Results

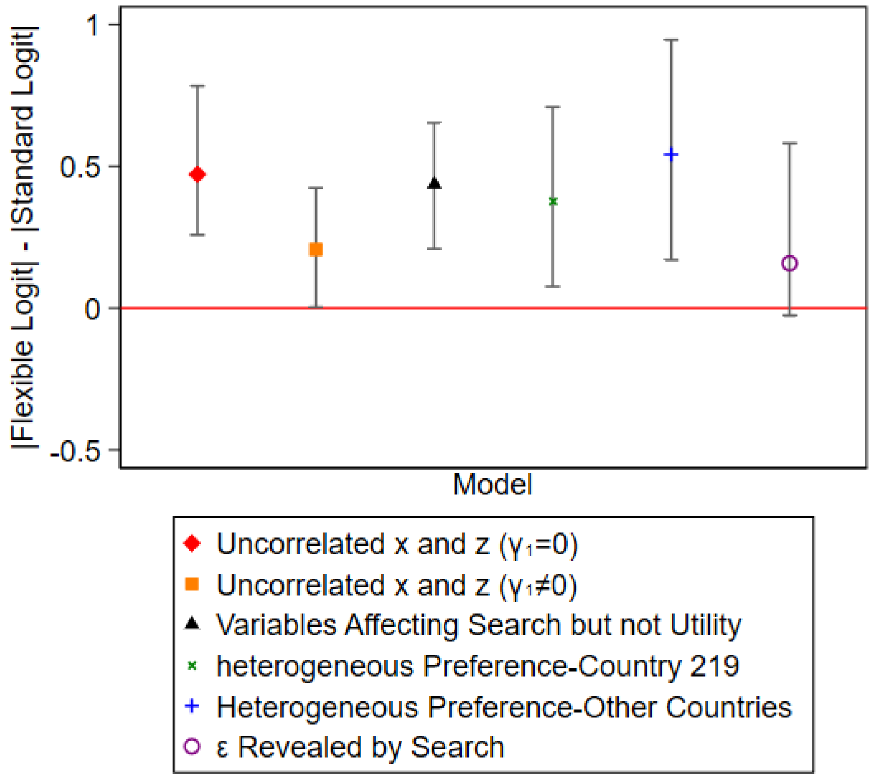
J.4.1 Alternative Model Specifications

To check the robustness of our results, we conduct a series of additional analyses for the case where location is the candidate z variable and the other attributes all serve as x variables. First, we estimate several extensions of our baseline model: (i) we allow consumers to form expectations about location based on the other attributes (i.e., $\gamma_1 \neq 0$); (ii) we let the position of the hotel on the results page affect the order of search as in Section 3.2; (iii) we consider the case where the idiosyncratic term ϵ_{ij} is only revealed to consumers after search (Section 3.3). As shown in Figure 5, we consistently find that the estimated preference for location in the standard logit model is attenuated compared with the flexible logit estimate. The only (slight) exception is for the case where ϵ is revealed via search, where we cannot reject the hypothesis that the two estimates are the same. We should note, however, that this version of the model is rejected by the data (we test the model by comparing ratios of first and ratios of second derivatives that the model implies should be equal, as detailed in Appendix A.6).

J.4.2 Bound Test on Expected Utility Assumption

Figure 6 shows the results of the bound test using the Expedia data with location score as the z variable. We can see that the bounds have some bite and the estimated choice probabilities almost always lie within the lower and upper bounds. These probabilities are thus broadly consistent with

Figure 5: Robustness of Results with Alternative Model Specifications



Note: This figure reports 95% confidence intervals for the difference between the absolute value of the normalized β estimate from flexible logit and the absolute value of the corresponding standard logit estimate across different models. For the "ϵ revealed by search" model, we report β/α_{price}.

the expected utility assumption.

J.4.3 Heterogeneous Preference

In Table 12 and Figure 7, we present summary statistics and histograms of the number of clicks conditional on searching good 1,2,...,10. They are not very different, suggesting that the preferences of consumers who searched good 1 are unlikely to be very different from those of the rest of the population. Furthermore, we utilize the demographic data on consumer country in the Expedia data, and split the full sample into two subsamples: Country 219 (32,859 choice sets, 60% of the data) and other countries. We estimate flexible logit separately in the two subsamples, and find attenuation in location preference consistent and stable. The difference test results are presented in the fourth and fifth specification in Figure 5.

J.4.4 Limited z Variation Method

In our identification argument, we recover the coefficients α on the attributes visible pre-search by conditioning on choice sets where all products have the same value of the z attribute. In practice,

Figure 6: Choice Probabilities, Upper and Lower Bounds from Expected Utility Assumption

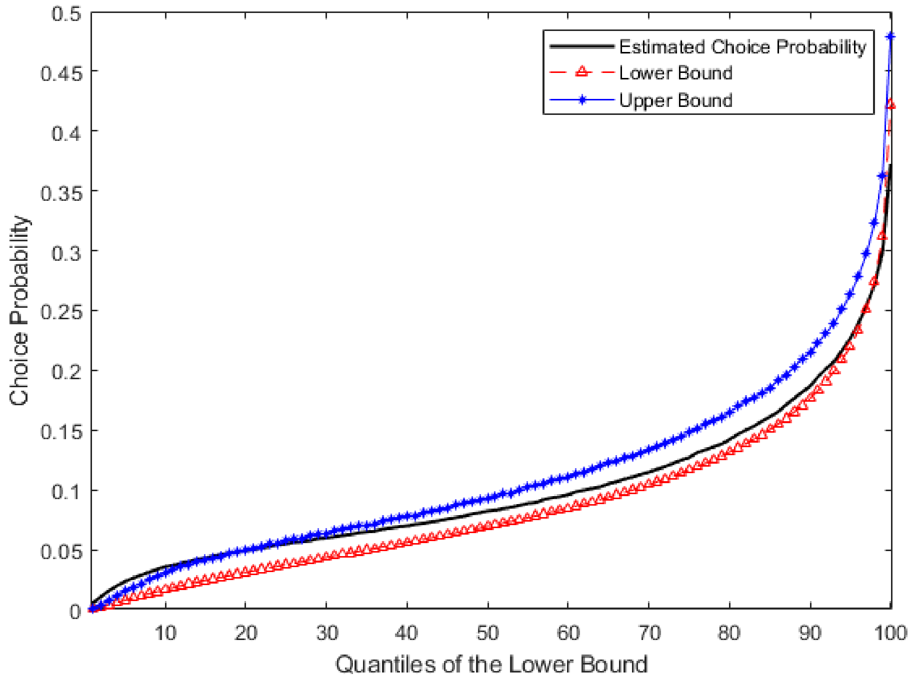
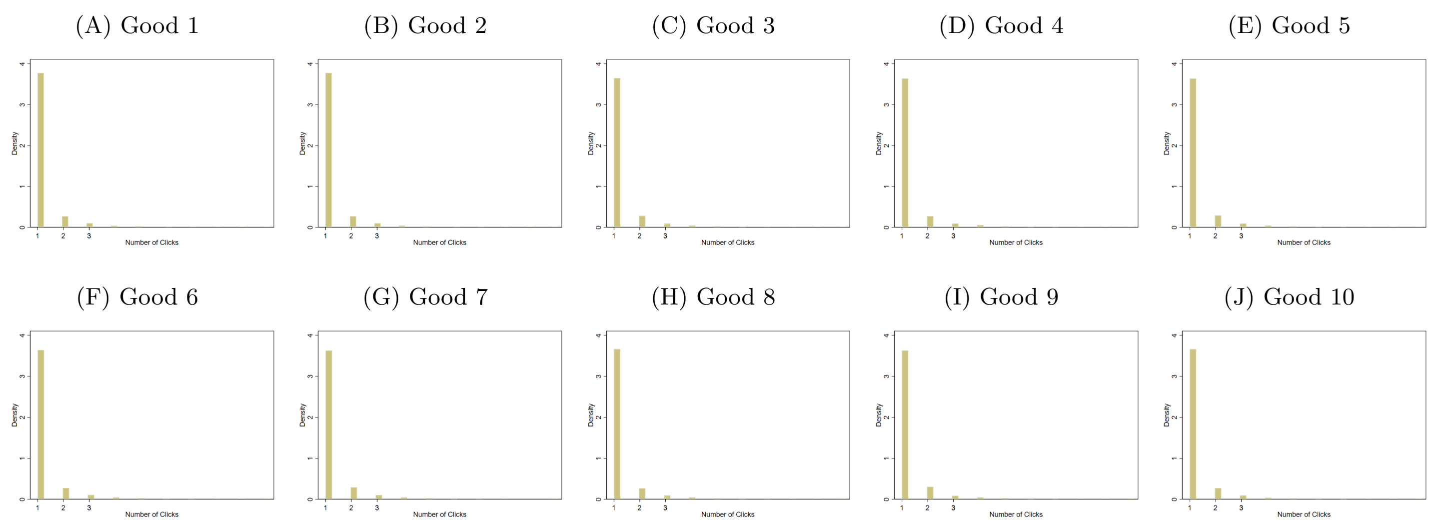


Figure 7: Total Number of Clicks Conditional on Searching Different Goods



Note: We report histograms of the total number of clicks conditional on consumer searching good 1,2,..., 10, where the goods are sorted in decreasing order of the hidden attribute z. There are no meaningful differences across the distribution of clicks, suggesting that consumers who searched the hotels with the best location did not exhibit substantially different search behavior than the rest of the users.

this entails taking a subsample in which z exhibits variation across products that is below a given threshold. It is then natural to ask how sensitive the results are to the choice of this threshold. To address this, we repeat our estimation using different thresholds for the z variation. Table 13 reports the results for both the lab experiment and the Expedia application. As expected, the estimate is

Table 12: Number of Clicks Conditional on Searching Different Products

Searched Good	Mean Number of Clicks	Standard Deviation	Number of Consumers
1	1.194	0.705	6,428
2	1.190	0.694	6,399
3	1.205	0.728	6,231
4	1.209	0.748	6,136
5	1.209	0.732	6,057
6	1.211	0.740	5,888
7	1.207	0.730	5,531
8	1.199	0.720	5,527
9	1.209	0.744	5,320
10	1.204	0.745	5,021

Note: We report key moments of the distribution of a user's total number of clicks conditional on the consumer searching good 1,2,..., 10, where the goods are sorted in decreasing order of hidden attribute z.

noisier if we use less of the data, but we find that our flexible estimates are reasonably stable and have consistently bigger magnitudes than the standard logit estimates.

J.4.5 β Estimates from Different Choices of x

Given that multiple variables can play the role of x in the Expedia data, we are free to choose any of them when taking derivatives to obtain our key ratio of second derivatives. If the model is correctly specified, the resulting estimates of the coefficient β on the location score should be statistically indistinguishable. Table 14 shows that this indeed is the case. Specifically, all confidence intervals overlap with the most accurate estimate obtained when price is chosen as the x variable.

Table 13: Robustness check on how much variation in z we use to recover α

	Point Estimate	Confidence Interval
Lab Experiment		
Standard Approach	-0.121	(-0.137, -0.105)
Flexible Approach		
Tercile	-0.275	(-0.383, -0.196)
Quartile	-0.272	(-0.367, -0.189)
Decile	-0.314	(-0.449, -0.167)
Ventile	-0.279	(-0.446, -0.110)
Percentile	-0.286	(-0.694, 0.066)
Expedia		
Standard Approach	0.298	(0.278, 0.317)
Flexible Approach		
Tercile	0.674	(0.495, 0.958)
Quartile	0.693	(0.513, 0.996)
Decile	0.770	(0.549, 1.108)
Ventile	0.869	(0.632, 1.276)
Percentile	0.800	(0.552, 1.169)

Note: We report point estimates and 95% confidence intervals for the β coefficients from different choices of how much variation in the z attribute we allow across goods when recovering the coefficient α. For instance, "Tercile" means that we only take the choice sets in which the standard deviation of z across products is in the bottom tercile. We find that the estimates from flexible logit are consistently stable, and significantly larger than the standard logit estimates.

Table 14: $\left(\alpha_k \cdot \left(\frac{\beta}{\alpha_k}\right)\right)$ estimates for different x variables

k	Point Estimate	Confidence Interval
Price	0.770	(0.549, 1.108)
Star	1.106	(-3.033, 14.975)
Review	0.647	(0.333, 7.118)

Note: We report point estimates and 95% confidence intervals for the β coefficients from different choices of the x variable when location score is treated as z. All confidence intervals overlap with price giving the most accurate estimate.