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

Jason Abaluck¶ Giovanni Compiani‡

November 8, 2020

Abstract

We state sufficient conditions under which choice data alone suffices to identify preferences when consumers are not fully informed about attributes of goods. Canonical models will be biased: the value of hidden attributes will be understated because consumers will be unresponsive to some variation in those attributes. In our baseline case, consumers search goods in order of the component of utility observable to them without search. Under our assumptions, an alternative method of recovering preferences using cross derivatives of choice probabilities succeeds under both full information and a range of search models and is thus robust to what consumers know when they choose. Our approach can be used to recover preferences from choices made by imperfectly informed consumers, to test for full information, and to forecast how consumers will respond to information. We verify in a lab experiment that our approach succeeds in forecasting the response to new information and assessing the value of that information when consumers engage in costly search.

*Thanks to Abi Adams, Judy Chevalier, Peter Hull, Magne Mogstad, Barry Nalebuff, 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, and Chicago for helpful comments and suggestions. Tianyu Han and Jaewon Lee 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`.

1 Introduction

When consumers purchase cars, houses, food, insurance, schooling and much else, they are often imperfectly informed about the attributes of relevant products in ways that substantially alter their choices (Allcott and Knittel 2019; Woodward and Hall 2012; Abaluck and Gruber 2011; Allcott, Lockwood, and Taubinsky 2019; Hastings and Weinstein 2008). Given this, models which assume full information may generate wrong conclusions about welfare and cannot be used to assess how choices would respond to more information. However, despite the emergence of behavioral economics as a major subfield of economic analysis, most work in applied economics continues to assume that choices are fully informed. 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.[1]

We believe this occurs for three reasons. First, for some positive purposes, it is irrelevant whether choices are informed since all that is required is to estimate how demand responds to price (Berry and Haile 2014). For instance, price elasticities are sufficient to predict equilibrium prices after a counterfactual merger between two firms.[2] Second, the data necessary to directly measure consumers' beliefs is often unavailable, and even when it is available, survey data is viewed with suspicion (Gul and Pesendorfer 2008). Full information is viewed as a parsimonious assumption in the absence of evidence to the contrary. Third, choice data alone does not suffice to separately identify preferences and beliefs without further assumptions (Manski 2002). Structural search models in which consumer beliefs can be identified (e.g., Ursu (2018)) require assumptions regarding whether consumers take into account option value, whether they solve an optimal stopping problem or "satisfice", distributional assumptions about prior beliefs and search costs, and whether choices are simultaneous or sequential, among others. The empirical literature suggests that canonical assumptions in all of these cases are often rejected by the data (respectively, Gabaix et al. (2006), Schwartz et al. (2002), Jindal and Aribarg (2018), Honka and Chintagunta (2016)), limiting the applicability of structural search models.

In this paper, we state what we believe is a more plausible condition under which choice data alone suffices to recover preferences whether consumers are fully or only partially informed. The approach relies on what we call *visible utility*, the component of utility visible to consumers prior to search (but not to the econometrician). In our baseline model, we impose that consumers search in decreasing order of visible utility (a condition we make precise below). We show that if this condition is satisfied, along with a few additional mild restrictions, 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 visible utility assumption. Specifically,

[1]The list of papers and their classification is available upon request from the authors.

[2]Even among the 126 articles in our survey that conduct welfare analyses and thus must take a stand on whether consumers are informed, 109 (86.5%) assume full information without testing this assumption.

no additional assumptions about option value, optimization vs. satisficing, simultaneous or sequential search, or distributional assumptions about beliefs and preferences are necessary for identification of preferences.

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.[3] As another example, our approach can inform firms' advertising strategies by, e.g., identifying product attributes that consumers care about but might not be currently aware of. Further, in settings where one would otherwise assume full information, the visible utility assumption provides a generalization which allows for both full and partial information, thus permitting a more realistic normative evaluation of choices. For example, Allcott, Lockwood, and Taubinsky (2019) propose taxing sugar-sweetened beverages to promote long-term health. A cost of this proposal might be that these foods are more desirable on other dimensions (e.g., tastiness), and conventional models would imply this if consumers appear willing to pay for high calorie foods. Allowing for imperfect information might reveal that consumers *prefer* low calorie foods once they are informed (e.g., for their physical appearance). In this case, the policy would be a win-win rather than one where health benefits must be weighed against short-term tastes. Finally, given preferences recovered by our approach, we show that it is possible to identify other primitives of interest under a maintained model of search. For example, in a sequential search model à la Weitzman (1979), one can identify the distribution of search costs and use them to simulate how information acquisition (and welfare) will vary across different contexts.

One can think of our approach as a data-driven method of isolating consumers who maximize utility. Consider the example of consumers purchasing items in a grocery store: nutritional information is accessible, but at some cost. Consumers may fail to maximize utility if they do not pay the cost to examine labels. In this case, visible utility represents utility from all non-nutrient sources, e.g. a combination of prices and perceived taste. Our assumption states that if you bother to check the nutrition label for good j, you will first check the label for a good j' that you would otherwise prefer were it equally nutritious. This assumption implies that consumers who search the most nutritious good *always choose the good that maximizes utility among all options* (which is not necessarily the most nutritious good). To see this, note that if some other good has higher utility than the most nutritious good, it must have higher visible utility and thus is searched and then chosen by the consumer. Further, only consumers who search the most nutritious good are sensitive to nutrient content for that good. Therefore, by looking at the sensitivity of choices to the nutrient content of the most nutritious good we are able to isolate consumers that behave *as if* they were fully informed; standard arguments then recover their preferences.

To spell things out in more detail, consider first a J-good model with linear utility $U_{ij} = x_j\alpha +$

[3] If an information intervention also reduces search costs, then the welfare gains via better choices given by our approach can be viewed as a lower bound to the total increase in welfare.

$z_j\beta + \epsilon_{ij}$ where $\alpha > 0$ and $\beta > 0$.[4] In the text, we extend this result to allow vector-valued x_j and z_j, as well as random coefficients and nonlinear utility. Suppose that consumer i observes x_j and ϵ_{ij} for all goods, but needs to engage in search to observe z_j. On the other hand, the researcher observes x_j, z_j, and choice probabilities s_j, but not ϵ_{ij}. With full information, we have $s_j = P(U_{ij} \geq U_{ij'}\ \forall j' \neq j)$ and we could estimate marginal rates of substitution using $\frac{\partial s_j}{\partial z_j} / \frac{\partial s_j}{\partial x_j} = \beta/\alpha$; in other words, β/α is identified by whether the choice probability for good j is more sensitive to z_j or x_j. If the underlying model is a search model in which consumers are informed about z_j only for some alternatives, then the standard approach will suffer from attenuation bias: $\left|\frac{\partial s_j}{\partial z_j} / \frac{\partial s_j}{\partial x_j}\right| < |\beta/\alpha|$. Some consumers will be insensitive to z_j variation not because they don't value it, but because they are not aware of it; thus, the observed sensitivity of choices to z_j will understate consumers' valuation of z_j. For each individual i and good j, we define *visible utility* as $VU_{ij} \equiv x_j\alpha + \epsilon_{ij}$. We call this quantity "visible utility" because it defines the utility that i receives from good j based only on x_j and ϵ_{ij}, the attributes of goods that consumers can observe without engaging in search (our main result also holds when ϵ_{ij} is only visible to consumers conditional on search, and so is not part of visible utility). In our baseline case, we assume that consumers search in decreasing order of visible utility. In other words, consumers always search first the goods that look more desirable given the information available to them. Note that the econometrician cannot tell the order of search since we do not observe ϵ_{ij}; this implies that observationally identical consumers can search in different orders.

The visible utility assumption is consistent with a broad class of search models. For example, in a Weitzman (1979) search model where the priors and search costs are the same across goods (but the latter vary across consumers), it is optimal to search the good in decreasing order of visible utility and decide whether to search the next good by comparing the expected benefits with search costs. Alternatively, consumers may myopically decide whether to continue searching by comparing utility in hand with expected utility of the next good (the "directed cognition" model of Gabaix, Laibson, Moloche, and Weinberg (2006)), consumers might engage in "satisficing," i.e. searching in order of visible utility and stopping whenever utility in hand is good enough, or they might simultaneously search all goods with visible utility above a certain threshold and then give up. In many cases, the underlying search process is simply unknown; in these cases, the conventional approach is to assume full information (potentially leading to biased estimates). Our approach is more general, allowing for full information, as well as a range of partial information models.

Our main result (for the case with linear utility and no random coefficients) is that, under the visible utility assumption and other conditions we make precise below, $\frac{\partial^2 s_1}{\partial z_1 \partial z_j} / \frac{\partial^2 s_1}{\partial z_1 \partial x_j} = \beta/\alpha$ for $j \neq 1$, where good 1 is defined as the good with the largest value of the hidden attribute z (which, again, is known to the econometrician but not necessarily to the consumer). This expression holds for any models where consumers search according to our assumptions above, including the full information

[4]We will show that the signs of α and β are identified, so assuming they are positive is without loss.

case; it is thus robust to whether consumers are fully or partially informed under our assumptions. The main downside of our approach relative to the full information assumption is that it is more demanding of the data, but this may be tolerable in large datasets typical of modern empirical work. The intuition for our result parallels the nutrition label example above: the expression $\frac{\partial s_1}{\partial z_1}$ is only non-zero for consumers who maximize utility, and consumers who maximize utility respond to attributes of rival goods (z_j and x_j) in proportion to their preferences. α is also separately identified from choice data alone,[5] and so our approach fully identifies preferences, not just marginal rates of substitution, and can be used for welfare analysis.

How general is this result? Using additional derivatives of the choice probability function, we can recover nonlinear utility functions $U_{ij} = v(x_j, z_j) + \epsilon_{ij}$. Additionally, the approach extends to random coefficients on product characteristics. Specifically, letting $U_{ij} = x_j\alpha_i + z_j\beta_i + \epsilon_{ij}$, we can recover the distribution of random coefficients (α_i, β_i) over a known grid. With a sufficiently long panel and time-invariant preferences, $U_{ijt} = v_i(x_j, z_j) + \epsilon_{ijt}$, we can recover individual-specific, possibly nonlinear utility functions $v_i(x_j, z_j)$. Thus, we can allow for a similar degree of unobserved heterogeneity as other constructive results on discrete choice demand with full information.[6] We also consider cases where the visible utility assumption is *not* satisfied, such as models with search costs varying across goods. We extend our model to allow for cases where (i) search costs vary with observables (e.g., rank on a webpage), (ii) consumers form expectations about z based on x, (iii) search reveals unobservable information, and (iv) either x or z is endogenous and valid instruments are available.

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. We also suggest an alternative parametric approach to estimate cross-derivatives that works well in simulations for larger numbers of goods. Given estimates of choice probabilities, 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 the attribute z is uncovered only after searching a good. That hypothesis can be tested provided that the data contains attributes x that can be assumed to be part of visible utility. Additionally, our model is overidentified;

[5]When z_j is the same for all goods, consumers maximize utility (although they themselves do not necessarily know this), so conventional methods suffice to recover α.

[6]Fox and Gandhi (2016) provide identification results for more general models allowing for both nonlinearity and flexible heterogeneity but these results are non-constructive and assume utility maximization; we use their result to identify parameters in corner cases where consumers maximize utility, but they otherwise are difficult to adapt to the more general case where choice probabilities need not maximize utility. This is in contrast to the constructive methods in Fox, Kim, Ryan, and Bajari (2012), who recover distributions satisfying the "Carleman condition," which implies that the distribution of preferences is uniquely characterized by its moments. Alternatively, we recover weights for distributions supported on a known and fixed grid, in line with the approach of Fox, Kim, Ryan, and Bajari (2011). Berry and Haile (2009) and Berry and Haile (2014) also provide related results on nonparametric demand estimation. Their focus is on recovery of the conditional distribution of utilities rather than the structural parameters of utility; the latter are essential for our task of assessing whether consumers are informed about relevant attributes.

in the case of homogeneous, linear preferences, for example, $\frac{\partial^2 s_1}{\partial z_1 \partial z_j} / \frac{\partial^2 s_1}{\partial z_1 \partial x_j}$ will be equal for all goods $j \neq 1$. We also show that the assumptions in our model imply nontrivial bounds on choice probabilities that can be checked in the data.

To validate our approach, we attempt to recover preferences in a lab experiment where individuals engage in costly search. Individuals choose from sets of three books with visible titles, authors, genre, star ratings and prices, but hidden discounts that can only be observed at some cost. We place no restrictions on which goods consumers search given these constraints. In particular, we do not constrain search to satisfy the visible utility assumption. For each individual, we also observe treatments where consumers choose given full information. As expected, conventional logit estimates using the costly search data give attenuated coefficients on the discount variable relative to the full-information case. By contrast, our search-robust estimates successfully recover full-information preferences. We show that our model successfully predicts the impact of an information intervention and permits an accurate welfare evaluation *before the intervention is conducted*. Estimated choice probabilities also satisfy bounds implied by the visible utility assumption.

Our result relates to several existing literatures. A large theoretical and empirical literature investigates the formation of "consideration sets".[7] This paper considers the complementary problem of imperfect information at the level of attributes rather than goods.[8] A few papers, such as Mehta, Rajiv, and Srinivasan (2003), Honka and Chintagunta (2016) and Ursu (2018) consider estimating utility by specifying full search models. We are, as far as we know, the first to provide formal identification results for preferences for a class of models without the need to commit to a specific structural search model.[9,10] A second related literature attempts to analyze 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

[7]Roberts and Lattin (1991), Conlon and Mortimer (2013), Goeree (2008), Gaynor, Propper, and Seiler (2016), and Barseghyan, Coughlin, Molinari, and Teitelbaum (2020). These papers attempt to estimate preferences when consumers may only consider some alternatives. Manzini and Mariotti (2014) and Abaluck and Adams (2017) are particularly closely related in attempting to characterize when we can recover consideration probabilities for alternative goods using choice data alone.

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

[9]There is one special case where the problem of imperfect information about attributes has been 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.

[10]An assumption in our model is that the searched attributes are observable to the econometrician, even if they are not known without searching to consumers. Ericson, Kircher, Spinnewijn, and Starc (2015) consider the related problem of inferring risk preferences separately from risk types using insurance choices. Their model differs from ours in that, in the special case they consider, the covariate "risk type" is not observed by the econometrician either.

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.

Section 2 lays out our formal framework and proves our identification results, Section 3 considers several empirically important extensions such as endogenous attributes, Section 4 provides details of estimation and simulation results, Section 5 discusses the (counterfactual) questions that can be addressed using our approach, Section 6 reports results from our experiment, and Section 7 concludes.

2 Model

There are $J \geq 2$ goods indexed by $j = 1, ..., J$ with attributes x_j observed by consumers and the econometrician and attribute z_j observed by the econometrician but not necessarily by consumers.[11,12] We assume that x_j and z_j are continuously distributed. For simplicity, we let x_j be scalar for all j, but our results immediately extend to the case of vector-valued x_j's. For now, we also focus on the case where z_j is a scalar and we let good 1 be the good with the largest value of z_j.[13,14]

Let individual i's utility from alternative j be denoted by $U_{ij}(x_j, z_j)$. In what follows, we often omit the dependence of U_{ij} on (x_j, z_j) unless it is necessary to avoid confusion. We can always write: $U_{ij} = a_{ij}(x_j) + b_{ij}(x_j, z_j)$ where $b_{ij}(x_j, 0) = 0$ (to see this, define $b_{ij}(x_j, z_j) = U_{ij}(x_j, z_j) - U_{ij}(x_j, 0)$). Since in our setting $a_{ij}(x_j)$ is the component of utility that is known to the consumer before engaging in search, we label it "visible utility," VU_{ij}. We make the following assumptions on the utility function.

Assumption 1. (i) For all i and j, U_{ij} is strictly monotonic in z_j.

(ii) For all i, the function $b_{ij}(x_j, z_j)$ is not alternative-specific, i.e. $b_{ij}(x_j, z_j) = b_i(x_j, z_j)$ for all j, and continuous in its first argument.

The class of utility functions satisfying Assumption 1 is broad and subsumes most specifications commonly used in empirical work as special cases, including logit with possibly nonlinear-in-characteristics utilities[15] and mixed-logit. For instance, in a mixed-logit model, one may specify $U_{ij} = \alpha_i x_j + \beta_i z_j + \epsilon_{ij}$. To map this specification into our notation, let $a_{ij}(x_j) = \alpha_i x_j + \epsilon_{ij}$, and

[11] 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.

[12] Since our model only requires variation in x and z for two goods, any of the remaining $J - 2$ goods may be taken to be the outside option.

[13] If there are multiple hidden attributes for each good, our results immediately apply if the data contains a choice set in which a good is more preferable than all other goods on each of the z attributes. Appendix A.2 shows that this can be relaxed in the homogeneous linear utility case.

[14] The definition of good 1 requires ruling out ties among the top two values of z_j.

[15] We allow for nonlinearities subject to Assumption 1(i) being satisfied.

$b_i(x_j, z_j) = \beta_i z_j$. As another example, consider the logit specification $U_{ij} = \alpha x_j + \beta z_j + \gamma x_j z_j + \epsilon_{ij}$. This is subsumed in our notation by letting $a_{ij}(x_j) = \alpha x_j + \epsilon_{ij}$, and $b_i(x_j, z_j) = \beta z_j + \gamma x_j z_j$.

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 VU_{ij}.

(ii) Conditional on having utility $\bar{u}$ in hand, consumer i searches j if and only if $g_i(x_j, \epsilon_{ij}, \bar{u}) \geq 0$ where g_i is decreasing in $\bar{u}$.[16]

(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 2.5. To briefly clarify, Assumption 2(i) states that consumers search goods in order of the component of utility visible to them without search (although not entirely visible to the econometrician). We view this as the strongest restriction in our model; Section 3 considers two relaxations that are relevant for empirical work. Assumption 2(ii) states that consumers decide whether or not to search a good based on their utility in hand and the visible utility of the good they are considering searching. 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. Further, Assumption 2(ii) also accommodates simultaneous search models in which consumers decide which goods to uncover based on visible utilities and then proceed to jointly search them. In this case, utility in hand is not a well-defined object and the function g_i does not vary with its second argument. 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) simply states that consumers must search a good before choosing it. Assumption 2(iv)—implicit in the model already—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 utility takes the form $U_{ij} = x_j\alpha_i + z_j\beta_i + \epsilon_{ij}$, 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 has the same prior F_z for all goods. Then, following Weitzman (1979), the consumer will rank goods according to their reservation value rv'_{ij} defined implicitly by*

$$c_i = \int_{rv'_{ij}}^{\infty} \left(u - rv'_{ij}\right) dF_{U_{ij}}(u) = \int_{rv_i}^{\infty} \beta_i \left(t - rv_i\right) dF_z\left(t\right) \tag{1}$$

[16] 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, we extend our model to permit search costs to vary across goods with observable factors.

where $rv_i \equiv \frac{rv'_{ij} - \alpha_i x_j - \epsilon_{ij}}{\beta_i}$ *and the last steps follows from a change of variable. We can interpret* rv_i *as the reservation value in units of* z*. To see this, note that consumer* i *ranks goods according to the visible utility* $x_j\alpha_i + \epsilon_{ij}$ *and for each good* j' *she chooses to uncover* $z_{j'}$ *if and only if the maximum utility secured so far is lower than* $x_{j'}\alpha_i + rv_i\beta_i + \epsilon_{ij'}$*. Once she stops searching, she maximizes utility among the searched goods. Thus, Assumption 2 is satisfied with* $g_i(x_j, \epsilon_{ij}, \bar{u}) = x_j\alpha_i + rv_i\beta_i + \epsilon_{ij} - \bar{u}$*.*

Example 2 (Directed Cognition Model)**.** *Suppose that utility takes the form* $U_{ij} = x_j\alpha_i + z_j\beta_i + \epsilon_{ij}$*. Further, as in the model of Gabaix, Laibson, Moloche, and Weinberg (2006), consumers rank goods in terms of expected utility*[17] *and myopically check whether searching the next good is worth the cost. The directed cognition model has the same* g_i *function as the Weitzman model,*[18] *but the order of search (and which goods are ultimately searched) may differ.*

Example 3 (Satisficing)**.** *Suppose that consumer* i *searches in order of visible 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 utility takes the form* $U_{ij} = x_j\alpha_i + z_j\beta_i + \epsilon_{ij}$ *and that consumer* i *simultaneously searches all goods that have visible utility above a threshold* $\tilde{\tau}_i$*. Then, Assumption 2 is satisfied with* $g_i(x_j, \epsilon_{ij}, \bar{u}) = \alpha x_j + \epsilon_{ij} - \tilde{\tau}_i$*.*

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 stand on each of these dimensions.

Throughout the rest of the paper, we assume without loss that $\frac{\partial b_i}{\partial z_j} > 0$, i.e. we treat z_j as an attribute that customers value in good j.[19] We are now ready to state and prove a lemma that is at the core of our results.

Lemma 1. *Let Assumptions 1 and 2 hold and let* $x_j \in [\bar{x} - \eta, \bar{x} + \eta]$ *for all* j*, for some* $\eta > 0$ *sufficiently small. If consumer* i *searches good 1 (i.e. the good with the highest value of* z*), then* i *chooses the utility-maximizing good.*

[17] Note that we may assume without loss that $E(z_j) = 0$ for all j since the mean value of the hidden attribute (known by rational consumers before search) is subsumed by visible utility.

[18] The result that consumers in the fully rational Weitzman model decide whether to continue searching "as if" they were myopic is one of the main insights of Weitzman (1979).

[19] Conditional on Assumption 1(i), 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. Thus, by the chain rule, the sign of $\frac{\partial b_i}{\partial z_j}$ 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 $z_1 > z_j$, by monotonicity, $b_i(\bar{x}, z_1) > b_i(\bar{x}, z_j)$. By continuity of b_i in its first argument, this implies that for η sufficiently small, $b_i(x_1, z_1) \geq b_i(x_j, z_j)$.[20] Given this, $U_{ij} > U_{i1}$ implies $VU_{ij} > VU_{i1}$. But by Assumption 2(*i*), this implies that good j is searched, which is a contradiction. □

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. When utility is linear, the same result holds under weaker conditions on the variation in x across products, which we formalize next.[21]

Lemma 2. *Let $U_{ij} = x_j\alpha_i + z_j\beta_i + \epsilon_{ij}$ and let Assumption 2 hold. If consumer i searches good 1 (i.e. the good with the highest value of z), then i chooses the utility-maximizing good.*

Proof. If good 1 is searched but utility is not maximized, then for some unsearched j, $U_{ij} > U_{i1}$. Since $z_1 > z_k$ for all $k \neq 1$, it must be that $x_j\alpha_i + \epsilon_{ij} > x_1\alpha_i + \epsilon_{i1}$. But by Assumption 2(*i*), this implies that good j is searched, which is a contradiction. □

Lemmas 1 and 2 will have far-reaching implications. To understand them, 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. Therefore, s_j is a function of $\mathbf{x} \equiv [x_1, \cdots, x_J]$ and $\mathbf{z} \equiv [z_1, \cdots, z_J]$, but we will often omit the dependence from the notation. Throughout the paper, the sources of unobserved heterogeneity will vary with the specific models we consider, so the symbol P will denote integrals over different distributions depending on the context.

Now, Lemmas 1 and 2 imply 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

[20] More formally, by continuity, for all $\delta > 0$ there exists $\eta > 0$ such that if $|x_1 - x_j| < 2\eta$, then $b_i(x_j, z_j) - b_i(x_1, z_j) \leq \delta$. Therefore, we have:

$$\begin{aligned} b_i(x_j, z_j) &= b_i(x_1, z_j) + b_i(x_j, z_j) - b_i(x_1, z_j) \\ &\leq b_i(x_1, z_j) + \delta \\ &\leq b_i(x_1, z_1) \end{aligned}$$

where the last inequality follows by choosing $\delta \equiv \frac{b_i(x_1, z_1) - b_i(x_1, z_j)}{2}$.

[21] Specifically, the restrictive condition on the support of x_j is only required for nonparametric identification, where strong support assumptions are often necessary.

preferences using standard arguments. To formalize this, note that Lemmas 1 and 2 imply we can write:

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

In other words, the probability that good 1 is chosen is the probability that good 1 is utility-maximizing minus the probability that good 1 is not searched *even though it is utility-maximizing.*[22] Failing to search good 1 requires that there exists some other good j with $VU_{ij} \geq VU_{i1}$ and utility high enough that $g_i(x_1, \epsilon_{i1}, U_{ij}) \leq 0$.

Our proof will use equation (3) and the fact that certain derivatives of the choice probability functions are linear in the preference parameters we hope to recover with known (or recoverable) weights. We consider identification for three specifications of utility that satisfy Assumption 1:

1. Cross-sectional data where $U_{ij} = v(x_j, z_j) + \epsilon_{ij}$
2. Panel data where $U_{ijt} = v_i(x_{jt}, z_{jt}) + \epsilon_{ijt}$
3. Cross-sectional data where $U_{ij} = x_j\alpha_i + z_j\beta_i + \epsilon_{ij}$

These cases are comparable in generality to existing constructive identification results for preferences in full information discrete choice models, such as Fox, Kim, Ryan, and Bajari (2012). In each of these cases, we assume that the unobservables are independent of the product characteristics $(\mathbf{x}, \mathbf{z})$. We will extend the approach to deal with endogenous attributes in Section 3.1.

2.1 Case 1: Cross-sectional data with $U_{ij} = v(x_j, z_j) + \epsilon_{ij}$

We start from the case where utility takes the form $U_{ij} = v(x_j, z_j) + \epsilon_{ij}$ for an unknown function v. In what follows, we use x and z to denote generic arguments of v and let $\epsilon_i \equiv [\epsilon_{i1}, \ldots, \epsilon_{iJ}]$.

Theorem 1. *Let Assumption 2 hold and utility be given by $U_{ij} = v(x_j, z_j) + \epsilon_{ij}$ with v increasing in both arguments and infinitely differentiable. Further, assume 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 $j^* \neq 1$, s_1 is infinitely differentiable and $\epsilon_i \perp (\mathbf{x}, \mathbf{z})$. Then, v is identified up to an additive constant.*

Proof. See Appendix A.1. □

This theorem applies to a broad class of utility functions. The cost of this level of generality is that it requires the share function s_1 to be infinitely differentiable. However, the marginal rates of substitution are recovered under much weaker differentiability requirements.

[22] This follows because Lemmas 1 and 2 imply that if good 1 chosen (and thus searched), good 1 must maximize utility.

Corollary 1. *Let Assumption 2 hold and utility be given by* $U_{ij} = v(x_j, z_j) + \epsilon_{ij}$ *with* v *increasing and differentiable in both arguments. Further, assume that* s_1 *is twice differentiable and* $\epsilon_i \perp (\mathbf{x}, \mathbf{z})$ *and assume that* $v(\cdot)$ *is identifiable from fully informed choices. Then, the marginal effects* $\frac{\partial v}{\partial z}$ *and* $\frac{\partial v}{\partial x}$ *can be identified using only second derivatives. Specifically, (i) marginal rates of substitution,* $\frac{\partial v}{\partial z} / \frac{\partial v}{\partial x}$*, can be recovered using:*

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

for all $j \neq 1$ *such that* $\frac{\partial^2 s_1}{\partial z_1 \partial x_j}(\mathbf{x}, \mathbf{z}) \neq 0$*, and (ii)* $\frac{\partial v}{\partial x}$ *can be identified from choices where* $z_j = z$ *for all* j*.*

We postpone the proof of Theorem 1 to Appendix A.1. Here, we focus on the special case with linear utility, which is very common in empirical work. In this case, the preference parameters can be fully recovered by looking at second derivatives. Further, to simplify the proof and facilitate intuition, we consider the setting with $J = 2$ goods. The result that second derivatives are sufficient to fully identify preferences immediately extends to the case with $J \geq 2$ goods (as we show in Appendix A.1, equation (24)).

Lemma 3. *Let utility be given by* $U_{ij} = x_j\alpha + z_j\beta + \epsilon_{ij}$ *and let Assumption 2 hold. Further, assume that* $\frac{\partial^2 s_1}{\partial z_1 \partial x_2}(\mathbf{x}^*, \mathbf{z}^*) \neq 0$ *for some* $(\mathbf{x}^*, \mathbf{z}^*)$*,* s_1 *is twice differentiable and* $\epsilon_i \perp (\mathbf{x}, \mathbf{z})$*. Then,*

$$\frac{\partial^2 s_1}{\partial z_1 \partial z_2}(\mathbf{x}^*, \mathbf{z}^*) \Big/ \frac{\partial^2 s_1}{\partial z_1 \partial x_2}(\mathbf{x}^*, \mathbf{z}^*) = \frac{\beta}{\alpha} \tag{5}$$

In addition, α *is identified by focusing on markets with* $z_j = z$ *for all* j *and thus* β *is also identified.*

Proof. First, we prove equation (5). In order to ease notation, we often suppress the subscript i in what follows. As above, good 1 is defined as the good with the highest value of z_j. Further, we let $\beta > 0$ without loss.[23] Using Lemma 2, the probability of choosing good 1 can be written as:

$$s_1 = P(U_1 > U_2) - P\left(\{U_1 > U_2\} \cap \{\mathcal{G} = 2\}\right) \tag{6}$$

where, as above, $\mathcal{G}$ denotes the set of searched goods. This follows because (i) if good 1 is utility-maximizing, you will always choose it unless you search only good 2; and (ii) you only choose good 1 if it is utility-maximizing, since otherwise, good 2 must have higher visible utility, meaning it must be searched (and chosen) if good 1 is searched.

Let $\tilde{u}_j \equiv x_j\alpha + z_j\beta$, 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

[23] This is without loss, since the sign of β is immediately identified from the data (footnote 19).

$\frac{\partial^2 s_1}{\partial z_1 \partial x_2} = \frac{\partial^2 s_1}{\partial z_1 \partial \tilde{u}_2} \frac{\partial \tilde{u}_2}{\partial x_2}$, and the result in equation (5) follows. To establish this, note that we can write:

$$\begin{aligned} P\left(\{U_1 > U_2\} \cap \{\mathcal{G} = 2\}\right) = \\ P(\{U_1 > U_2\} \cap \{VU_2 > VU_1\} \cap \{g(x_1, \epsilon_1, U_2) \le 0\}) = \\ P(\{U_1 > U_2\} \cap \{g_1(x_1, \epsilon_1, U_2) \le 0\}) - P(\{VU_1 > VU_2\} \cap \{g_1(x_1, \epsilon_1, U_2) \le 0\}) \end{aligned} \tag{7}$$

where the second line follows since $VU_1 > VU_2$ implies $U_1 > U_2$ and thus $P\left(\{VU_1 > VU_2\} \cap \{g_1(x_1, U_2) \le 0\}\right) = P\left(\{U_1 > U_2\} \cap \{VU_1 > VU_2\} \cap \{g_1(x_1, U_2) \le 0\}\right)$. The second term on the last line of display (7) 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, together with equation (6), is sufficient to show 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 $\tilde{u}_2$, thus proving equation (5).

Finally, we show that we can identify α using standard techniques by looking at choice sets where $z_j = z$ for all j. To see this, note that when $z_j = z$ for all j then consumers maximize utility if and only if they maximize visible utility. Since by assumption they always search the good with the highest visible utility, it follows that they maximize utility. Thus, one can pin down α by looking at how the choice probabilities vary with $\mathbf{x}$ conditional on $z_j = z$ for all j, just like in the full information case.[24] Given (5) and α, identification of β follows immediately. □

Finally, we note that in many models of interest the conventional way of identifying preferences based on the ratio of *first* derivatives leads to understating consumers' taste for z. For this result, we further assume that the function $g_i(x_j, \epsilon_{ij}, \bar{u})$ is weakly increasing in x_j.[25] For simplicity, consider the model with linear utility $U_{ij} = x_j\alpha + z_j\beta + \epsilon_{ij} \equiv \tilde{u}_j + \epsilon_{ij}$ and $\alpha > 0$. Let $\tilde{\mathbf{u}} \equiv (\tilde{u}_1, \tilde{u}_2)$ and

$$\begin{aligned} P^*_{j,2}\left(\tilde{\mathbf{u}}, \mathbf{x}\right) &\equiv P(\{U_j > U_{-j}\} \cap \{VU_{-j} > VU_j\} \cap \{g_i(x_j, \epsilon_{ij}, U_{-j}) \le 0\}) \\ P^*_{j,3}\left(\tilde{\mathbf{u}}, \mathbf{x}\right) &\equiv P(\{U_{-j} > U_j\} \cap \{VU_j > VU_{-j}\} \cap \{g_i(x_{-j}, \epsilon_{i-j}, U_j) \le 0\}) \end{aligned}$$

Then, $s_j = P(U_j > U_{-j}) - P^*_{j,2}\left(\tilde{\mathbf{u}}, \mathbf{x}\right) + P^*_{j,3}\left(\tilde{\mathbf{u}}, \mathbf{x}\right)$. As we have shown above, for good 1, $P^*_{1,3}\left(\tilde{\mathbf{u}}, \mathbf{x}\right) = 0$, but this is not necessarily true for other goods.

[24] There is one subtle exception to this argument. Suppose there is an outside option with utility normalized to 0, and we wish to identify a fixed effect which gives the utility of all inside goods relative to the outside good. In this case, consumers do not necessarily maximize utility when $z_j = z$ for all goods because consumers may decide to search *none* of the inside goods, and they may do so even when the outside good has lower utility than some of the inside goods if search costs are sufficiently high (in other words, an outside option may violate our assumption that consumers must search a good before they choose it). When consumers search none of the inside goods, it is never possible to separately identify whether consumers do not value the inside goods or have high search costs to examine any of the inside goods. It is possible to say something about the utility of consumers who are induced to search at least one of the inside goods when price is low enough (for example), but parametric assumptions are needed to make claims about the utility of consumers who never search any of the inside goods.

[25] This condition 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.

Differentiating, we obtain:

$$\begin{aligned}
\frac{\partial s_j}{\partial z_j} &= \beta\left[\frac{\partial P(U_j > U_{-j})}{\partial \tilde{u}_j} - \frac{\partial P^*_{j,2}}{\partial \tilde{u}_j}(\tilde{\mathbf{u}}, \mathbf{x}) + \frac{\partial P^*_{j,3}}{\partial \tilde{u}_j}(\tilde{\mathbf{u}}, \mathbf{x})\right] \\
\frac{\partial s_j}{\partial x_j} &= \alpha\left[\frac{\partial P(U_j > U_{-j})}{\partial \tilde{u}_j} - \frac{\partial P^*_{j,2}}{\partial \tilde{u}_j}(\tilde{\mathbf{u}}, \mathbf{x}) + \frac{\partial P^*_{j,3}}{\partial \tilde{u}_j}(\tilde{\mathbf{u}}, \mathbf{x}) - \frac{1}{\alpha}\frac{\partial P^*_{j,2}}{\partial x_j}(\tilde{\mathbf{u}}, \mathbf{x}) + \frac{1}{\alpha}\frac{\partial P^*_{j,3}}{\partial x_j}(\tilde{\mathbf{u}}, \mathbf{x})\right]
\end{aligned}$$

Note that $\frac{\partial P(U_j > U_{-j})}{\partial \tilde{u}_j} - \frac{\partial P^*_{j,2}}{\partial \tilde{u}_j}(\tilde{\mathbf{u}}, \mathbf{x}) + \frac{\partial P^*_{j,3}}{\partial \tilde{u}_j}(\tilde{\mathbf{u}}, \mathbf{x}) = \frac{\partial s_j}{\partial \tilde{u}_j} \geq 0.$[26] Further, $\frac{1}{\alpha}\frac{\partial P^*_{j,2}}{\partial x_j}(\tilde{\mathbf{u}}, \mathbf{x}) \leq 0$ and $\frac{1}{\alpha}\frac{\partial P^*_{j,3}}{\partial x_j}(\tilde{\mathbf{u}}, \mathbf{x}) \geq 0$ due to our assumptions about the function g. Therefore,

$$\frac{\frac{\partial s_j}{\partial z_j}}{\frac{\partial s_j}{\partial x_j}} \leq \frac{\beta}{\alpha} \tag{8}$$

This shows that standard discrete choice models that assume full information—such as multinomial logit or probit—will typically suffer from attenuation bias under our assumptions.

2.2 Case 2: Panel data where $U_{ijt} = v_i(x_{jt}, z_{jt}) + \epsilon_{ijt}$

This case closely parallels the proof in the previous section. Now, rather than observing only $s_j(\mathbf{x}, \mathbf{z})$, the choice probabilities for each alternative as a function of the attributes, panel data allows us to observe $s_{ij}(\mathbf{x}, \mathbf{z})$, the choice probabilities for each individual as $(\mathbf{x}, \mathbf{z})$ vary over a long period of time. Given these, the following result holds:

Theorem 2. *Let Assumption 2 hold and utility be given by $U_{ijt} = v_i(x_{jt}, z_{jt}) + \epsilon_{ijt}$ with v_i increasing in both arguments and infinitely differentiable. Further, assume that $\frac{\partial^2 s_{i1}}{\partial z_1 \partial x_{j^*}}(\mathbf{x}^*, \mathbf{z}^*) \neq 0$ for some $(\mathbf{x}^*, \mathbf{z}^*)$ and $j^* \neq 1$, s_{i1} is infinitely differentiable and $\epsilon_i \perp (\mathbf{x}, \mathbf{z})$. Then, v_i is identified up to an additive constant.*

Corollary 2. *Let Assumption 2 hold and utility be given by $U_{ijt} = v_i(x_{jt}, z_{jt}) + \epsilon_{ijt}$ with v_i increasing and differentiable in both arguments. Further, assume that s_{i1} is twice differentiable, $\epsilon_i \perp (\mathbf{x}, \mathbf{z})$, and assume that $v(\cdot)$ is identifiable from fully informed choices. Then, the marginal effects $\frac{\partial v_i}{\partial z}$ and $\frac{\partial v_i}{\partial x}$ can be identified using only second derivatives. Specifically: marginal rates of substitution, $\frac{\partial v_i}{\partial z} / \frac{\partial v_i}{\partial x}$, can be recovered using:*

$$\frac{\partial^2 s_{i1}}{\partial z_1 \partial z_j}(\mathbf{x}, \mathbf{z}) \Big/ \frac{\partial^2 s_{i1}}{\partial z_1 \partial x_j}(\mathbf{x}, \mathbf{z}) = \frac{\partial v_i}{\partial z}(x, z) \Big/ \frac{\partial v_i}{\partial x}(x, z)$$

for all $j \neq 1$ such that $\frac{\partial^2 s_{i1}}{\partial z_1 \partial x_j}(\mathbf{x}, \mathbf{z}) \neq 0$, and $\frac{\partial v_i}{\partial x}$ can be identified from choices where $z_j = z$ for all j.

---

[26] Increasing 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 u_j increases the probability good j is chosen. Second, changing u_j doesn't change the probability that good j is searched, which depends on $g_i(x_j, \epsilon_{ij}, U_{-j})$ for each alternative searched good. Third, changing u_j never makes other goods more likely to be searched. Specifically, an alternative good k is searched if and only if $g_i(x_k, \epsilon_{ik}, U_{k'}) \geq 0$ 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, $\frac{\partial s_j}{\partial \tilde{u}_j} \geq 0$.

The proofs of these two results exactly parallel the arguments in the previous section.

2.3 Case 3: Cross-sectional data where $U_{ij} = x_j\alpha_i + z_j\beta_i + \epsilon_{ij}$

The cases in the previous two sections assume either that we have panel data or that all individual heterogeneity is additively separable. Due to the difficulty of separately identifying preferences and search as well as more practical difficulties with estimation, most empirical structural search models that we are aware of do not allow for non-separable unobserved heterogeneity (see, e.g., Ursu (2018), Honka, Hortaçsu, and Vitorino (2017)).

Of course, we would like to understand both from a theoretical perspective whether the assumption of separable heterogeneity is required for identification and from a practical perspective whether our results are applicable in such cases. The canonical case of non-separable heterogeneity that has been studied in the literature and for which constructive identification results exist is that of the linear random coefficients model. We maintain linearity and impose two additional assumptions.

Assumption 3. (i) Utility is given by $U_{ij} = x_j\alpha_i + z_j\beta_i + \epsilon_{ij}$.

(ii) The coefficients α_i and β_i take values on a known finite support, i.e. $\alpha_i \in \{\alpha_1, \cdots, \alpha_{K_\alpha}\}$ and $\beta_i \in \left\{\beta_1, \cdots, \beta_{K_\beta}\right\}$ with probabilities given by $\tilde{\pi}_{k_\alpha,k_\beta} \equiv P\left(\{\alpha_i = \alpha_{k_\alpha}\} \cap \left\{\beta_i = \beta_{k_\beta}\right\}\right)$. Further, the elements of $\left\{\beta_1, \cdots, \beta_{K_\beta}\right\}$ all have the same sign and, without loss, we assume that they are positive.

(iii) The distribution of ϵ_i is known (or independently identified) and the three random vectors $\epsilon_i, (\alpha, \beta)$ and $(\mathbf{x}, \mathbf{z})$ are mutually independent.

Assumption 3(ii) follows Fox, Kim, Ryan, and Bajari (2011) and a recent strand of empirical papers (e.g., Nevo, Turner, and Williams (2016)) in assuming that the random coefficients are supported on a finite and known grid of points. Given the restriction that $\left\{\beta_1, \cdots, \beta_{K_\beta}\right\}$ all have the same sign, assuming that they are positive is without loss (see footnote 19). Assumption 3(iii) maintains knowledge of the distribution of all unobservables other than the random coefficients, consistent with recent papers on identification and estimation of demand (e.g. Fox, Kim, Ryan, and Bajari (2012), Fox, Kim, and Yang (2016)).

Theorem 3. *Let Assumptions 2 and 3 hold. If the market share of good 1 is $K_\alpha K_\beta-$time differentiable, then the probability weights $\tilde{\pi}_{k_\alpha,k_\beta}$ for $k_\alpha = 1, \cdots, K_\alpha$, $k_\beta = 1, \cdots, K_\beta$ are identified.*

Proof. See Appendix A.3. □

As it is clear from the statement of Theorem 3, allowing for heterogeneity across consumers in preferences for attributes typically requires taking derivatives of order higher than two. Thus, identifying heterogeneous preferences is more demanding of the data. When the sample size does not allow for direct application of our result, a natural approach is to impose more structure by specifying a structural search model. Theorem 3 may then be used to establish nonparametric identification of preferences within the specified model of search.[27]

Finally, we note that Theorem 3 focuses on recovering the entire distribution of the random coefficients. If the goal is simply to test whether consumers have full information, then taking second-order derivatives turns out to be sufficient under certain conditions, which we spell out in Appendix B.

2.4 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 visible utility for a subset of goods grows to infinity (minus infinity), the probability of searching those goods goes to one (zero) 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 z_j (or our weighted index in the vector-valued case, see Appendix A.2). To recover preferences in the homogeneous linear case, we only need sufficient variation to estimate second derivatives of s_1 at a single point. Of course, flexibly recovering a nonparametric function v or nonparametric random coefficients requires substantially more variation and data in order to estimate higher order derivatives of choice probabilities. We further discuss these challenges in Section 4.

2.5 Discussion of Search Model Assumptions

To reiterate, we consider search models satisfying the following assumptions:

1. Consumer i searches goods in decreasing order of VU_{ij}.

2. Conditional on having utility $\bar{u}$ in hand, consumer i searches j if and only if $g_i(x_j, \epsilon_j, \bar{u}) \geq 0$ where g_i is decreasing in $\bar{u}$.

3. Consumers choose the good which maximizes utility among searched goods.

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

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 visible attributes of those goods, the distribution of the hidden attribute z_j, and

[27]See Section 5 for more on this point.

search costs. If z_j is i.i.d. across goods and consumers have the same search cost for all goods, then it follows that consumers will search in order of visible utility (see Example 1). There are at least three reasons this might fail in the Weitzman (1979) model: 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 $VU_3 > VU_2$. Third, search costs might vary across goods, meaning that consumers prefer to search first goods with lower search costs even if the payoff is potentially lower.

While the restriction that priors be i.i.d. and search costs be constant across goods is sufficient for Assumption 2(i) (the first assumption above), this is not necessary. 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 visible 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)). In Section 3.2, we extend our main result to allow for these violations of our visible utility assumption by considering cases where some observable attributes impact search but not utility. We can also relax the i.i.d. priors assumption by allowing consumers to form beliefs about the hidden attribute as a function of observed attributes. Specifically, in Section 3.3, we extend our approach to the case where beliefs about z_j are a linear function of observables.

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 visible utility above a threshold, random search, and directed cognition. One exception is a model in which consumers simultaneously search the top K goods in terms of visible utility prior to engaging in search. 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 visible utility of all goods. We show in section 3.5 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 Section 4.

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, and the second is that they maximize utility given the information available. 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 assumption that consumers maximize utility given the information available can also be relaxed. One could specify 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. The first is that only the value of z_j is unknown prior to search. A consumer who clicks through to the product information page of an Amazon product might learn information about the attributes of a good ("the battery is compatible with USB-c"), but they also might learn information not observable to the econometrician ("one reviewer said the battery exploded into flames"). In section 3.4, we show that our results continue to hold if the ϵ component of utility is revealed only conditional on search (as in Kim, Albuquerque, and Bronnenberg (2010) and Ursu (2018)). The second piece of the fourth assumption is 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. This is not always plausible: 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), among others).

2.6 Testing Search Model Assumptions with and without Observable Search

Our proof 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 naively assume that our identification results would be unnecessary in such cases; given data on which products were searched, perhaps preferences can be estimated conditional on search 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. One example would be if search depends on ϵ. In such cases, goods with undesirable observables that are searched likely have an especially high realization of ϵ. Thus, it will appear from conditional choice probabilities 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. 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.[28] Thus, with heterogeneous preferences, the existing literature requires specifying a search model in order to estimate preferences even when search is observed. Our approach avoids the need to do this under the assumptions we have outlined.

Once our approach is used to identify preferences, clickstream data can be used to conduct additional overidentifying tests if we assume that the distribution of ϵ_{ij} is known. In the linear case, visible utility is given by $VU_{ij} = x_j\alpha_i + \epsilon_{ij}$. As shown in Lemma 3, examining choices with equal values of the hidden attribute is sufficient to identify the distribution of α_i. Given α_i, the known distribution of ϵ_{ij}, and the number of goods searched $|\mathcal{G}_i|$, we can thus compute:

$$P(j \in \mathcal{G}_i|\mathbf{x}, \mathbf{z}) = \sum_k P(|\mathcal{G}_i| = k|\mathbf{x}, \mathbf{z})P(j \in \mathcal{G}_i||\mathcal{G}_i| = k, \mathbf{x}, \mathbf{z}) \tag{9}$$

since the first probability on the RHS is observed and the second is pinned down by the model assumptions (specifically, the fact that with k goods searched, those k goods must be the k goods with the highest visible utility). Checking (9) 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 visible utility and higher utility. Thus, we have:

$$s_j(\mathbf{x}, \mathbf{z}) \leq 1 - P(U_{ik} \geq U_{ij} \text{ and } VU_{ik} \geq VU_{ij} \text{ for some } k) \tag{10}$$

The latter probability can be directly computed from knowledge of preferences and the distribution of ϵ. To construct a lower-bound, note that the probability of choosing good j is at least as large as the probability that good j maximizes both utility and visible utility. That is:

$$s_j(\mathbf{x}, \mathbf{z}) \geq P(U_{ij} \geq U_{ik} \text{ and } VU_{ij} \geq VU_{ik} \text{ for all } k) \tag{11}$$

[28] 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.

Once again, this probability can be computed given knowledge of preferences and the distribution of ϵ. We can then check whether our estimated choice probabilities are consistent with these bounds.

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

2.7 Testing for Full Information

Our results suggest a natural tests for full information. Consider first the case of homogeneous preferences as in Section 2.1. Under the null hypothesis of full information, $s_j = P(U_j \geq U_k \ \forall k)$ and therefore:

$$\frac{\partial^2 s_1}{\partial z_1 \partial z_j} \Big/ \frac{\partial^2 s_1}{\partial z_1 \partial x_j} = \frac{\partial s_k}{\partial z_j} \Big/ \frac{\partial s_k}{\partial x_j} = \frac{\partial v}{\partial z} \Big/ \frac{\partial v}{\partial x} \tag{12}$$

for all $j \neq 1$ and all k. On the contrary, when consumers are unaware of z_j for some goods, then the ratios of first derivatives need not be equal to the ratios of the second derivatives. For example, equation (8) showed that $\frac{\partial s_1}{\partial z_1} / \frac{\partial s_1}{\partial x_1} \leq \frac{\partial^2 s_1}{\partial z_1 \partial z_j} / \frac{\partial^2 s_1}{\partial z_1 \partial x_j}$ in the class of search models we consider. Since both the ratios of first derivatives and the ratios of second derivatives in (12) are estimable from the data, this immediately leads to a test based on the discrepancy between the two sets of ratios. More specifically, given estimators of the share functions, one can compute a Wald test-statistic based on the discrepancy between the two sets of ratios and reject the null hypothesis of full information if the statistic exceeds a critical value.

Note that this test is valid even if our assumptions on the search process fail to hold since with full information the two sets of ratios will be equal regardless. When our assumptions on the search process do hold, we expect the test to have power, since the first derivative ratio will be attenuated relative to the true preferences, which are recovered by the cross-derivative ratio.

In the case of heterogeneous preferences, the ratio of first derivatives might be attenuated relative to the ratio of second derivatives even under the null hypothesis of full information. In Appendix B, we provide verifiable sufficient conditions that rule this out and therefore guarantee the validity of our test for the mixed logit model with random coefficients distributed on a grid (as in Section 2.3).

3 Extensions

In this section, we consider several extensions to the baseline model. These extensions are designed to accommodate features prominent in the empirical search literature. More specifically, we consider:

- Endogenous attributes

- Attributes which impact search but not utility
- Expectations about z which depend on x
- Search reveals unobservables
- The "$K-$rank" simultaneous search model

In each of the extensions we discuss, the same cross-derivative expressions from our original proofs suffice to identify preferences, in some cases subject to additional assumptions.

We consider each extension separately, although in principle, one could estimate models combining several such extensions. For example, Kim, Albuquerque, and Bronnenberg (2010) estimate a search model in which only attributes unobservable to the econometrician are revealed during search, and in which some observables impact search but not utility.

3.1 Endogenous attributes

So far, we have assumed that the observed product attributes are independent of all unobservables. This is 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

$$U_{ij} = \alpha x_j + \beta_i z_j + \lambda_i p_j + \xi_j + \epsilon_{ij} \tag{13}$$

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.[29] If firms also know ξ_j when choosing p_j, then the two will typically be correlated, thus leading to endogeneity of p_j. We consider both the case where p_j is part of visible utility and that in which consumers need to search good j to uncover p_j (as well as possibly other non-endogenous attributes 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

[29] Note that the utility specification in (13) allows for random coefficients on both z_j and p_j, but not on x_j. This is stronger than needed, since the identification argument below only requires that x_j and ξ_j enter the demand functions via a linear index. Thus, another possible specification is

$$U_{ij} = \tilde{\alpha}_i \left(\alpha x_j + \xi_j\right) + \beta_i z_j + \lambda_i p_j + \epsilon_{ij}$$

The latter is weaker, but also less common in the discrete choice literature, so we focus on model (13) in what follows.

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 assumption 2(ii). Let $\delta_j = \alpha x_j + \xi_j$ for all j.

Assumption 4. (i) The attribute p_j is part of the visible utility of good j. Conditional on having utility $\bar{u}$ in hand, consumer i searches j if and only if $g_{ij}(\delta_j, p_j, \bar{u}) \geq 0$ where g_{ij} is decreasing in $\bar{u}$.

(ii) The attribute p_j is uncovered by consumers only upon searching good j. Conditional on having utility $\bar{u}$ in hand, consumer i searches j if and only if $g_{ij}(\delta_j, \bar{u}) \geq 0$ where g_{ij} is decreasing in $\bar{u}$.

Like Assumption 2(ii), Assumption 5 states that consumers decide whether to search good j based on utility in hand and the visible utility of j. In Appendix A.4, we invoke results from Berry and Haile (2014) to show that these assumptions suffice for nonparametric identification of the choice probability functions provided we have valid instruments (in a sense we make precise in the Appendix). Once the choice probability functions are identified, one may apply our results in Section 2.3 to identify the distribution of the preference parameters α, β_i and λ_i.

3.2 Allowing for variables affecting search but not utility

One important case in which the visible utility assumption 2(i) is likely to fail is when factors exist which 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 has no direct impact on utility conditional on searching (Ursu 2018). In this case, consumers might first search items with higher search position even if they do not have higher visible 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. A second example is if we observe advertising expenditures for each good and believe that advertising entices consumers to search advertised goods.

Our model from Section 2.1 can be extended to deal with cases where the factors which impact search but not utility are observable and the sign of their impact on search probabilities is known (such as position in search). Denote the variable which perturbs search but not utility by r_j, suppose that r_j is observed 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 VU_{ij} alone, we assume that goods are searched based on $m(VU_{ij}, r_j)$ where m is strictly increasing in both VU_{ij} and r_j. We show in Appendix A.5 that a version of our identification argument continues to hold provided we see sufficient variation in product attributes conditional on search position.

3.3 Allowing for consumers' expectations on z to depend on x

Another reason why the visible utility assumption 2(i) might fail is that consumers could form expectations about z based on x. For instance, if x is price and z is quality, consumers might infer that more expensive products tend to be higher quality. As a consequence, if they value quality to a sufficient degree relative to price, they may search a high-priced product and not search a low-priced product even if the former has a lower visible utility than the latter.

In our proofs so far, we have not made any explicit assumption about whether consumers update about z_j given x_j, but such updating is likely to lead to violations of the visible utility assumption if not explicitly modeled. We now show that we can identify preferences given consumer beliefs about z_j given x_j in a linear model. Further, under additional assumptions, we will show that we can identify β/α, the relative value of the hidden attribute, even when beliefs are unknown. In other words, we can do so without taking a stand on whether consumers have rational expectations and form beliefs based on the empirical relationship between z_j and x_j or naively update. Consider the linear model $U_{ij} = x_j\alpha + z_j\beta + \epsilon_{ij}$ and re-write it as

$$\begin{aligned} U_{ij} &= x_j\alpha + (z_j - E(z_j|x_j))\beta + E(z_j|x_j)\beta + \epsilon_{ij} \\ &= \beta\gamma_0 + x_j(\alpha + \beta\gamma_1) + \tilde{z}_j\beta + \epsilon_{ij} \end{aligned}$$

where the second equality assumes that consumers use the linear projection $E(z_j|x_j) = \gamma_0 + \gamma_1 x_j$ and we let $\tilde{z}_j \equiv z_j - E(z_j|x_j)$. Visible utility is then given by $\beta\gamma_0 + x_j(\alpha + \beta\gamma_1) + \epsilon_{ij}$ and consumers learn the deviation from their expectation on z_j, $\tilde{z}_j$, upon searching. Note that γ_0 is not identified, but also does not generally impact choices since it enters utility as an additive constant.[30]

In Appendix A.6, we show that given γ_1, we can recover β and α using an analog of our usual approach. When γ_1 is not observed, we can still identify β/α if we know its sign and assume that we observe goods with the largest value of z_j and the smallest value of x_j. The quantity β/α is not sufficient to simulate choices with full information, since we cannot tell how responsive consumers would be to x_j were choices fully-informed. However, it is sufficient to identify the relative value placed on the hidden attribute as well as to conduct tests for full information as in Section 2.7.

3.4 Unobservables revealed by search

So far, we have 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)). To accommodate this, we consider a modification of our model where the shock ϵ_{ij} only becomes known to consumer i upon searching good

[30] There is one exception to the above claim, which is the case when there is an outside option for which the x and z attributes are not defined, so that a systematic bias in beliefs about the distribution of z_j given x_j would change the relative value of all the inside goods relative to that outside option. This might mean that the relative utility of the outside option cannot be separately identified from γ_0; the model could still be estimated, but the normative interpretation of fixed effects might change.

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. This means that, in Assumption 2, VU_{ij} is now equal to αx_j and Assumption 2(iv) is dropped.

Given this setup, we show in Appendix A.7 that the ratio of second derivatives $\frac{\partial^2 s_j}{\partial z_j \partial z_k} \Big/ \frac{\partial^2 s_j}{\partial z_j x_k}$ recovers $\frac{\beta}{\alpha}$ provided that one chooses good k to be the good with the highest value of x (note that j need not be the good with the highest value of z here). Thus, our approach can be extended to deal with the possibility that search reveals unobservables.

3.5 The $K-$rank Simultaneous Search Model

As noted above, our main model allows for consumers to choose which goods to search in one simultaneous step. However, 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 visible utility (e.g., 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 visible utility of good j, but on the visible utility of all other goods as well, thus violating Assumption 2(ii).

In Appendix A.8, we show via an alternative argument that the usual second-derivative ratio from equation (5) still identifies $\frac{\beta}{\alpha}$ in the two-good $K-$rank model. We also show in our simulation results that our method succeeds in a model where consumers search the top K goods (with K varying randomly across consumers).

4 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. Note that the model 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.[31] 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.

[31] Here, we focus on the case where data on individual-level choices are available, as in the experiment of Section 6. However, our identification approach could also be applied to aggregate (i.e. market share) data as long as one can consistently estimate the share functions s_j.

Here, we consider two approaches to estimating $s_1(\mathbf{x}, \mathbf{z})$: (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. Note that a good in our model is defined by the collection of attributes observable to the econometrician (potentially including good fixed effects); in other words, different products with the same attributes count as the same good. Thus, estimation of choice probabilities and their derivatives does not require that all consumers have identical products in their choice sets, or even that the same products are available to many different consumers (unless product fixed effects are of interest). What we need is sufficient variation in attributes to flexibly estimate the mapping from the product attributes to choices.

Throughout this section, we focus on the linear homogeneous case of $U_{ij} = x_j\alpha + z_j\beta + \epsilon_{ij}$. Our result in Section 2.1 shows that β/α can be recovered from $\frac{\partial^2 s_1}{\partial z_1 \partial z_j} / \frac{\partial^2 s_1}{\partial z_1 \partial x_j}$ for $j \neq 1$. Relative to conventional estimation of linear homogeneous discrete choice models, our approach is more demanding of the data, requiring estimation of second derivatives for a specific good. Especially in large datasets where more flexible estimation is feasible, this allows us to be more agnostic about the underlying information structure.

As discussed in Section 2.3, the model with linear, homogeneous preferences is the current standard in the empirical literature on search (e.g. Mehta, Rajiv, and Srinivasan (2003), Honka and Chintagunta (2016) and Ursu (2018); Kim, Albuquerque, and Bronnenberg (2010) is a notable exception in that they allow for random coefficients). In more general non-linear or random coefficients models, our identification arguments require recovery of higher-order derivatives and thus might not directly translate into viable estimation strategies in small to medium sample sizes or with a large number of goods. In these cases, the best way forward might be to parametrically specify a full structural search model and estimate it via standard methods, e.g. MLE. We would then view our identification results as providing reassurance that preferences are indeed identified, something that had not been formally established in the literature (see Section 5 and Appendix D for more on this). Additionally, given the estimated structural search model, one can use the derivations in Section 2.6 to test the model restrictions (subject to the visible utility assumption).

4.1 Approximation via Bernstein polynomials

Following Compiani (2019), 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 depends on the attributes of the goods, but not their identity.[32] Exchangeability is

[32] See Compiani (2019) for a formal definition of exchangeability.

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 nonparametric 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. More details on the implementation of the estimator can be found in Compiani (2019).

Simulation 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:

$$U_{ij} = \alpha x_{ij} + \beta z_{ij} + \epsilon_{ij} \tag{15}$$

with $\alpha = \beta = 1$, $x_{ij} \sim_{i.i.d} N(0,1)$, $z_{ij} \sim_{i.i.d.} N(0,1)$, and ϵ_{ij} i.i.d. Type 1 extreme value.

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

1. Weitzman search, with search costs $c \sim LogNormal(-2, 2.25)$
2. Satisficing, searching in order of visible utility until utility-in-hand is at least $T \sim LogNormal(-0.35, 2.25)$
3. Search all goods with visible 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 visible utility)
4. Randomly search $K \in \{1, \ldots, J\}$ goods, where the searched goods are the K highest in terms of visible 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 visible utility, but on the visible utilities of all goods.

Table 1 reports results from the Bernstein approximation of the cross-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 2.1: 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 cross-derivative ratio include the true values in DGPs 1-3, and are fairly precise for the $J = 3$ case. For DGP 4, where the assumptions of our model do not hold (see Section 3.5), the coefficient is attenuated for $J = 3$, although the point estimates remain much closer to the true values the first-derivative estimates.

Table 1: Bernstein Approximation

	Number of Goods			
	2		3	
DGP	First-Derivatives	Cross-Derivatives	First-Derivatives	Cross-Derivatives
1	0.610	0.977	0.403	0.997
	(0.024)	(0.304)	(0.012)	(0.076)
2	0.691	1.280	0.361	0.935
	(0.024)	(0.538)	(0.014)	(0.068)
3	0.527	0.870	0.330	0.872
	(0.021)	(0.190)	(0.010)	(0.071)
4	0.444	0.801	0.206	0.626
	(0.018)	(0.301)	(0.010)	(0.075)

Note: Across all rows, the data 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 errors, obtained via 250 bootstrap repetitions, are reported in parentheses.

4.2 "Flexible Logit"

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

To motivate our parametric approach to estimating $s_1(\mathbf{x}, \mathbf{z})$, note that 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. extreme value where v_j^* is a differentiable function of x_j and z_j, then for $q_j \in \{x_j, z_j\}$:

$$\begin{aligned}
\frac{\partial s_j}{\partial q_j} &= \frac{\partial s_j}{\partial v_j^*}\frac{\partial v_j^*}{\partial q_j} = \frac{\partial v_j^*}{\partial q_j} s_j(1 - s_j) \\
\frac{\partial s_j}{\partial q_{j'}} &= \frac{\partial s_j}{\partial v_{j'}^*}\frac{\partial v_{j'}^*}{\partial q_{j'}} = -\frac{\partial v_{j'}^*}{\partial q_j'} s_j s_{j'} \\
\frac{\partial^2 s_j}{\partial z_j \partial q_{j'}} &= -\frac{\partial v_j^*}{\partial q_{j'}}\frac{\partial v_j^*}{\partial z_j} s_j s_{j'}(1 - 2s_j)
\end{aligned} \tag{16}$$

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 cross-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:

$$v_1 = \tilde{v}(x_1, z_1) + b_1 z_1 + \sum_{k \neq 1} (\gamma_k w_{z1k} z_k + \gamma_{2k} w_{x1k} x_k + w_{z2k} \delta_k z_k z_1 + w_{x2k} \delta_{2k} x_k z_1) \tag{17}$$

where $\tilde{v}(x, z)$ is a differentiable function of x and z, w_{z1k}, w_{x1k}, w_{z2k} and w_{x2k} are known weights, and b_1, γ_k, γ_{2k}, δ_k and δ_{2k} are coefficients to be estimated. Further, we let $v_k = \tilde{v}(x_k, z_k)$ for $k \neq 1$. In Appendix C, we describe one way of choosing the weights which we find works well in simulations, and for which the ratio of second derivatives (which recovers β/α) is a particularly convenient function of model parameters. We note that the parameters in (17) do not have the usual 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, (17) is simply a flexible function of $(\mathbf{x}, \mathbf{z})$ that captures the second derivatives of s_1 well.

Simulation Results For each of the DGPs described in Section 4.1, we consider simulations with $J \in \{2, 3, 5, 10\}$. We report estimates from the flexible logit model as well as the naive logit model. We bootstrap the standard errors using 250 repetitions.

Results from these simulations are reported in Table 2. The table shows estimates of β/α from a conditional logit model with no adjustment for imperfect information, as well as the cross-derivative ratio estimates from the flexible logit model. In the naive 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 and 5 goods, and have less bias than the naive logit model for 5 and 10 goods.

Table 2: Estimator based on cross-derivatives ratio (flexible logit) vs naive logit

	Number of Goods							
	2		3		5		10	
DGP	Naive	Flexible	Naive	Flexible	Naive	Flexible	Naive	Flexible
1	0.6590	1.0214	0.6330	1.0671	0.6050	0.9633	0.5770	0.8986
	(0.0158)	(0.1208)	(0.0122)	(0.1259)	(0.0095)	(0.1254)	(0.0089)	(0.1053)
2	0.7403	0.9976	0.6194	1.0854	0.4587	1.0407	0.2909	1.0004
	(0.0162)	(0.1034)	(0.0135)	(0.1300)	(0.0102)	(0.1578)	(0.0083)	(0.2603)
3	0.5424	1.1177	0.5945	1.0286	0.6543	0.9017	0.7246	0.8822
	(0.0149)	(0.1716)	(0.0117)	(0.1469)	(0.0099)	(0.1071)	(0.0106)	(0.0733)
4	0.4543	1.1358	0.5568	0.9614	0.6691	0.8015	0.7887	0.8151
	(0.0140)	(0.1906)	(0.0118)	(0.1659)	(0.0105)	(0.1012)	(0.0104)	(0.0679)

Note: Across all rows, the data the sample size is $N = 20,000$ and the data in each row is generated by the corresponding DGP described in the main text. "Naive" 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 errors, obtained via 250 bootstrap repetitions, are reported in parentheses

5 Counterfactuals

Our results can be used to identify preferences under the assumptions we outline without estimating a full search model. In this section, we discuss for which counterfactuals preference estimation is sufficient. Additionally, we discuss how our results can be used to aid in estimation of search costs given a fully specified search model or for specification testing after such a search model is estimated.

5.1 Applications without Recovering Search Costs

Benefits of Full Information 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.[33] In these cases, knowing preferences is sufficient to simulate how information would impact choices without a structural search model. For example, 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 and value of interventions before they are conducted. Of course, our method quantifies the welfare gains from more informed choices, but not the gains directly stemming from reduced search costs. In this sense,

[33]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 3.4), 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.

the estimated increase in welfare can be viewed as a lower bound 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 (see the next subsection) or using some auxiliary data on, e.g., time spent searching and value of time.

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., Bagwell (2007), Becker and Murphy (1993)).

Allocation Problems A third important class of counterfactuals involves direct reallocation. For example, employers who offer multiple health insurance plans may consider offering only a single plan to some consumers (Bhargava, Loewenstein, and Sydnor 2017). Preferences over factors such as customer service ratings or the scope of provider networks may be relevant (Handel and Kolstad 2015). To evaluate whether direct reallocation improves welfare via the quality of the resulting option, knowing preferences is sufficient. No search model is needed.

Normative Evaluation of Choices In many counterfactuals where limited information or search costs are not the primary object of interest, one nonetheless is concerned to accurately value attributes of goods. In the introduction, we give the example of a tax on sugar-sweetened beverages. An alternative 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 consumers to the products they would otherwise value if they had more information, meaning that it is 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.

5.2 Applications with Search Costs

We focused above on applications where search costs do not need to be recovered. Our model can also be used to identify search costs given preferences and an underlying structural search model. In Appendix D, 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. There are several reasons search costs might be of interest.

Welfare with Structural Search Costs 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 value of making choices more easily. 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 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. 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 often be more credible and less prone to misspecification than structural estimates (e.g. Kling, Mullainathan, Shafir, Vermeulen, and Wrobel (2008)).

Counterfactuals with Non-Zero Search Costs 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. Modeling explicitly how these changes would impact search costs for different goods, and thus which goods are chosen, would require an explicit search model.

Validating Parametric Models A final reason to estimate a full structural search model is to impose parametric restrictions on the data necessary for estimation in finite samples. Our identification proof shows that, in principle, these parametric restrictions are unnecessary for identification. This is confirmed by the simulation results we presented for models with linear utility and homogeneous preferences over observables. However, when the coefficients on attributes are heterogeneous—something the empirical search literature typically rules out—estimation of the higher-order derivatives of choice probabilities necessary for nonparametric identification (see Section 2.3) may not be possible given the data available. In such cases, a natural approach is to specify a structural search model with random coefficients in order to place some parametric structure on these higher-order derivatives. This requires taking an explicit stand on the underlying search model. Nonetheless, once the model has been estimated and preferences recovered, the results in Section 2.6 can be used to conduct specification tests. If these tests reject, an alternative search model may fit the data better.

6 Experimental Validation

Our identification proof and simulation results show that preferences can be estimated regardless of whether consumers are fully informed, provided the underlying model is a search model satisfying the stated assumptions. Of course, the theorem 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 the assumptions of our model. We nonetheless show that we are able to correctly recover preferences using our "search-robust" estimation technique.

6.1 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.

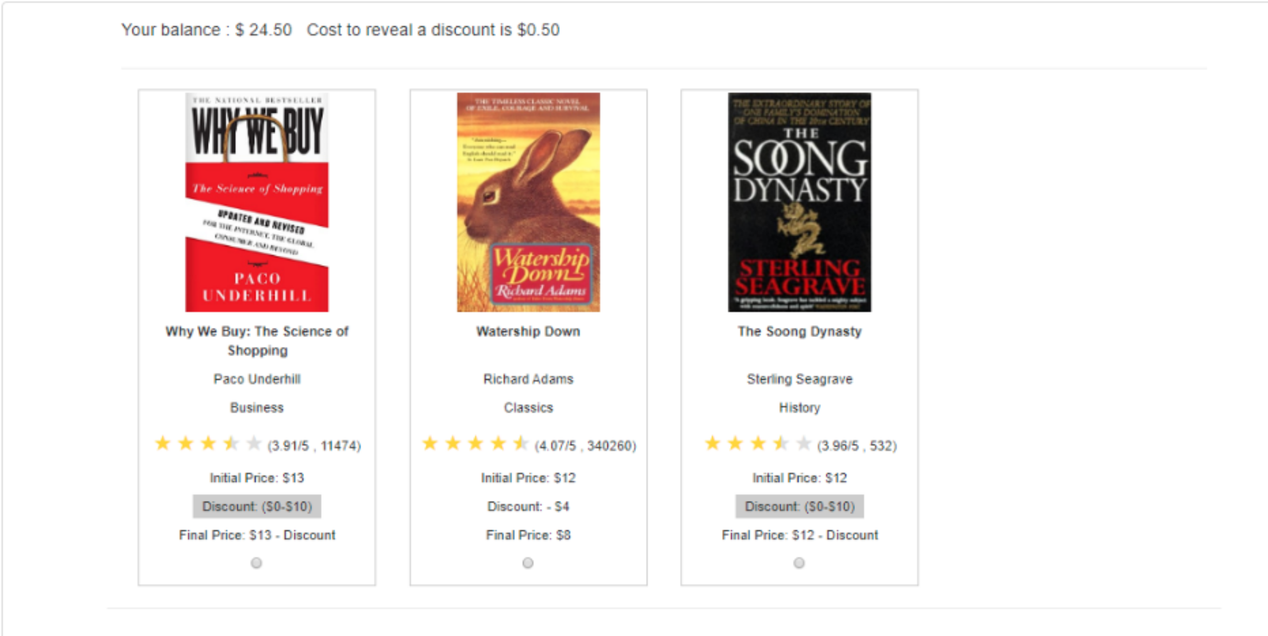
In our experiment, conducted via Mechanical Turk, each participant made 40 choices from sets of 3 randomly selected books. For all books, participants could see a photo of the cover, the title, author and genre, as well as the Goodreads rating and the number of ratings. Prices were randomized to integers from \$11-\$15 (equally likely). All books were then further discounted by an integer amount from \$0-\$10 (equally likely). All users were given a \$25.00 bank at the start of each choice, from which any costs incurred were deducted. There were a total of 93 participants, yielding 3,720 choices.

The discount is our key variable of interest. For 10 of the 40 choices, users could see all discounts and thus could see the net price of all options at no cost. For 30 of the 40 choices, discounts were hidden and users had to pay a cost to see the discount for any given book.[34] The cost per click was constant for each user across the 30 choices, and randomly chosen from $\{\$0.10, \$0.25, \$0.35, \$0.50\}$. For the 30 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 40 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 \$25.00.

Figure 1 shows a sample product selection screen from a choice where discounts were hidden. In this 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 10 choices where all information is revealed are our benchmark for the "truth." The goal is then to test whether the relative weight on discounts and prices that we estimate in the cases where discounts are costly to observe matches the relative weight we see when discounts

[34] The full information and costly information choice situations were randomly ordered, so that the 10 "full information" choices were intermixed with the costly information choices.

Figure 1: Lab Experiment: Sample Product Selection Screen



are visible to everyone. Further, because both discounts and prices are in dollar terms, and because they are randomized (and so not signals of quality), there is a second benchmark: if consumers are rational, the weight on discounts and prices should be equal.

In other words, we will model choices using:

$$U_{ij} = price_j \cdot \alpha_1 - discount_j \cdot \beta + rating_j \cdot \alpha_2 + \epsilon_{ij} \tag{18}$$

where ϵ_{ij} is i.i.d. type-I extreme value[35] and accounts for any aspects of consumers taste 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.

6.2 Estimation Results

Columns 1 and 2 of Table 3 show results from estimating a conditional logit model on consumer choices for the 10 choice situations (per consumer) where all information is revealed (Full Info) and the 30 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

[35] We will compare nonparametric estimates based on our approach to estimates from a conventional logit model. Only the latter requires distributional assumptions on ϵ_{ij}.

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 conditional 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 0.986 in the full information treatment and 0.683 in the costly information treatment.

Table 3: Conditional Logit and Cross-Derivative Estimation Results

Variable	Full Info	Costly Info	Cross-Derivative Ratio
Price	-0.386***	-0.302***	-0.387***
	(0.038)	(0.018)	(0.032)
Discount (-)	-0.376***	-0.206***	-0.399
	(0.020)	(0.009)	-
Rating	0.591***	0.421***	0.584***
	(0.190)	(0.099)	(0.161)
Discount (-) / Price	0.986***	0.683***	1.032***
	(0.093)	(0.044)	(0.102)
N	930	2790	2790

Note: The table shows estimation results from a conditional logit model estimated on the full information and costly information treatments in columns 1 and 2, and estimation of the cross-derivative ratio based on Bernstein polynomials in column 3. The minus sign indicates that discount multiplied by -1 so that the coefficient on discount should equal that of price with full information. Standard errors on 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%.

Following Section 4.1, we estimate the demand function $s_1(\mathbf{x}, \mathbf{z})$ via Bernstein polynomials. Specifically, we use the tensor product of univariate Bernstein polynomials, one for each argument of the s_1 function.[36] Further, we impose the natural constraint that s_1 be decreasing in the price of book 1 and the discount of books 2 and 3, and increasing in the discount of book 1 and the price 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.[37] The estimate is 1.032, which is close to the corresponding number from column 1. In addition to estimating β/α_1, we need to directly recover the α coefficients. Consistent with Lemma 3, we compute these by estimating a conditional logit model using only choice sets where the variance

[36] 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 21.

[37] 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.

of the discount across goods is in the bottom quintile. The results are reported in column 3 of Table 3, along with the value of β implied by our estimates of α_1 and β/α_1. 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. Further, the confidence interval is sufficiently tight to exclude the logit estimates in the costly information treatment.

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 \$0.66 per choice from choosing books with lower discounts. The corresponding number in the data is \$0.69 per choice situation, since consumers in the costly information treatment average discounts of \$6.24, while consumers in the full information treatment average discounts of \$6.93. 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. 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 \$0.15 based on column 3. Thus, our model again yields results that are quite close to those coming from the "true" fully informed choices in the data.[38]

6.3 Testing the Visible Utility Assumption

As in most real-world settings, visible 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. The assumption that consumers search according to the visible utility assumption 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 consumers reveal with full information.

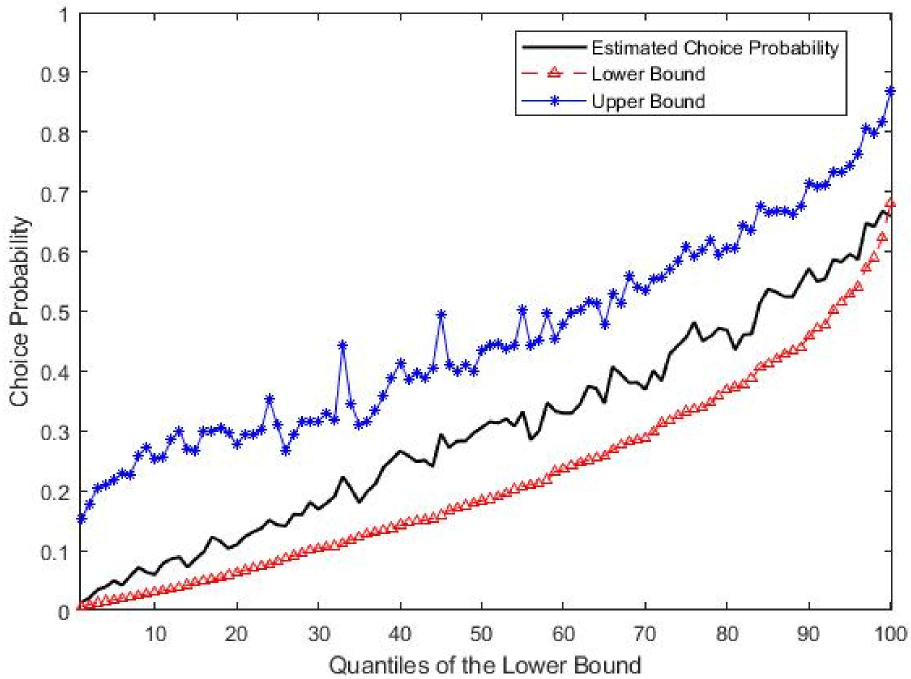
As discussed in Section 2.6, while the visible utility assumption cannot be verified directly, it can be tested along with the other restrictions of our model. One such test is to compute bounds on the choice probabilities implied by the model. Given our estimates of preferences and assumptions about the distribution of ϵ_{ij}, we can compute the upper and lower bounds described in Section 2.6 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

[38]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.

estimated via Bernstein polynomials.

Figure 2 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 15 to 30 percentage points. The estimated choice probabilities in nearly all cases lie within this range. These probabilities thus appear broadly consistent with the visible utility assumption.

Figure 2: Choice Probabilities, Upper and Lower Bounds from Visible Utility Assumption



7 Conclusion

We prove that it is possible 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. This result holds in a broad class of search models. The functions of choice probabilities which identify preferences in our model are "robust" in the sense that they work in both full information and search models. 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. Prior to conducting an information intervention, choice data can be used to estimate counterfactually how consumers would choose were they fully informed. If preferences are not informed, the preferences consumers would have if they were informed can be used to conduct a more defensible welfare analyses.

Preferences can (sometimes) be identified in structural search models, but such models require making many explicit assumptions about how consumers search. Do consumers consider option value or are they myopic? Do they solve an optimal stopping problem or search until they find a good enough option? Is search sequential or simultaneous? If search costs vary across consumers, what is their statistical distribution? Our approach attempts to avoid these complexities by instead relying on a sufficient condition satisfied in a broad class of search models that can be falsified by the available data via bounds on choice probabilities and overidentification tests.

In many settings, one can conceive of reasons that the visible utility assumption would fail, but it 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. Even if one lacks contextual information to support the visible utility assumption, our approach is much weaker than the standard assumption of full information and may be preferable in settings where preferences are needed to conduct welfare analysis. The main downside of our approach relative to full information is statistical power, but this concern is less relevant given rich microdata which is increasingly available. In settings where one would otherwise make many untested structural assumptions about search, visible utility may be more parsimonious and leads to clear, testable predictions.

Our assumptions are sufficient for identification but not necessary. This raises several questions for future research: are there other conditions aside from the visible 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?[39]

Increasingly, empirical analyses relax the assumption that consumers make informed choices. Typically, behavioral welfare analysis is done using auxiliary data, restrictions on preferences, or by testing whether consumers choose differently when provided with information. Despite this, absent data to the contrary, the default assumption in most economic analysis remains that consumers make informed choices. Our result suggests this need not be the case. Even with no auxiliary data, researchers can use observed choices both to test whether choices are informed and to recover what preferences would be were consumers more informed. This removes the (often compelling) excuse that while consumers may not be informed, assuming informed choices is the only constructive way to proceed given the data available.

[39] This question parallels Falmagne (1978)'s derivation of necessary and sufficient conditions for choice probabilities to be rationalizable by utility maximization given full information. Our question differs by relaxing the full information assumption.

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*.

Allcott, H. and C. Knittel (2019). Are consumers poorly informed about fuel economy? evidence from two experiments. *American Economic Journal: Economic Policy 11*(1), 1–37.

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.

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. Teitelbaum (2020). Heterogeneous consideration sets and preferences.

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 (2009). Nonparametric identification of multinomial choice demand models with heterogeneous consumers. Technical report, National Bureau of Economic Research.

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

Bhargava, S., G. Loewenstein, and J. Sydnor (2017). Choose to lose: Health plan choices from a menu with dominated option. *The Quarterly Journal of Economics 132*(3), 1319–1372.

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.

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

Compiani, G. (2019). Market counterfactuals and the specification of multi-product demand: a nonparametric approach. *Working Paper*.

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.

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.

Ericson, K. M., P. Kircher, J. Spinnewijn, and A. Starc (2015). Inferring risk perceptions and preferences using choice from insurance menus: theory and evidence. Technical report, National Bureau of Economic Research.

Falmagne, J.-C. (1978). A representation theorem for finite random scale systems. *Journal of Mathematical Psychology 18*(1), 52–72.

Fox, J. T. and A. Gandhi (2016). Nonparametric identification and estimation of random coefficients in multinomial choice models. *The RAND Journal of Economics 47*(1), 118–139.

Fox, J. T., K. i. Kim, S. P. Ryan, and P. Bajari (2011). A simple estimator for the distribution of random coefficients. *Quantitative Economics 2*, 381–418.

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.

Fox, J. T., K. i. Kim, and C. Yang (2016). A simple nonparametric approach to estimating the distribution of random coefficients in structural models. *Journal of Econometrics 195*(2), 236–254.

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.

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.

Gul, F. and W. Pesendorfer (2008). The case for mindless economics. *The foundations of positive and normative economics: A handbook 1*, 3–42.

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 (2016). 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.

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.

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

Manski, C. F. (2002). Identification of decision rules in experiments on simple games of proposal and response. *European Economic Review 46*(4-5), 880–891.

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

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.

Nevo, A., J. L. Turner, and J. W. Williams (2016). Usage-Based Pricing and Demand for Residential Broadband. *Econometrica 84*(2), 411–443.

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.

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

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.

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 use the notational convention $\frac{\partial f}{\partial x^0}(x) = f(x)\,\forall x$ for any function f.

A.1 Proof of Theorem 1

We let $2 = j^*$ and $(\mathbf{0}, \mathbf{0}) = (\mathbf{x}^*, \mathbf{z}^*)$ for notational convenience (the proof is unchanged for other values of j^* and $(\mathbf{x}^*, \mathbf{z}^*)$).

We show that the derivatives of the share functions identify the derivatives of v at a point and thus the entire function up to a constant. Fix any (x, z) in the domain of v and consider a Taylor expansion of v around the point (0,0):

$$v(x,z) = v_0 + \frac{\partial v_0}{\partial x}x + \frac{\partial v_0}{\partial z}z + ... + \frac{1}{n!}\frac{\partial^n v_0}{\partial z^{\bar{n}} \partial x^{n-\bar{n}}} z^{\bar{n}} x^{n-\bar{n}} + ... \tag{19}$$

where $v_0 \equiv v(0,0)$ and $\frac{\partial^n v_0}{\partial z^{\bar{n}} \partial x^{n-\bar{n}}} \equiv \frac{\partial^n v}{\partial z^{\bar{n}} \partial x^{n-\bar{n}}}(0,0)$ for all $n \geq 1, \overline{n} \leq n$. To recover $v(x,z)$, it is sufficient to recover all derivatives $\frac{\partial^n v_0}{\partial z^{\bar{n}} \partial x^{n-\bar{n}}}$. First, we map this into the notation of Assumption 1 as follows:

$$U_{ij} = \underbrace{v(x_j, 0) + \epsilon_{ij}}_{a_i(x_j)} + \underbrace{v(x_j, z_j) - v(x_j, 0)}_{b(x_j, z_j)}$$

Note that we can directly recover $v(x,0)$ by the following argument. When $z = 0$, $b(x,0) = 0$ which means that consumers who search the highest visible utility good (which is guaranteed by Assumption 2(i)) maximize utility and we can identify $v(x,0)$, a function of x only, using standard techniques. By differentiating $v(x,0)$ and evaluating at $x = 0$, we can recover all of the terms that contain no z's, i.e. $\frac{\partial^n v}{\partial x^n}(0,0)$.

To recover derivatives of v with respect to z, we will use Lemma 1. Specifically, we will take $x_j \in [\bar{x} - \eta, \bar{x} + \eta]$ for all j, where $\bar{x}$ and η are defined in Lemma 1, and use the fact that $\frac{\partial s_1}{\partial z_1}$ can be written as a function of terms which only depend on x_2 and z_2 via U_2. To formalize this, we let $\mathcal{J}_1 \equiv \{2, \ldots, J\}$, $v_j \equiv v(x_j, z_j)$ for all j, and $\mathbf{v} \equiv (v_1, \ldots, v_J)$. Similarly, we let $v_j^0 = v(x_j, 0)$ and $\mathbf{v}^0 = \left(v_1^0, \ldots, v_J^0\right)$. Then, by (3) we can write for all $(\mathbf{x}, \mathbf{z})$ with $z_1 \geq z_j$ for all j:

$$\begin{aligned}
s_1 = P(U_1 \geq U_k \ \forall k) - &\sum_{\mathcal{S} \subset \mathcal{J}_1, \mathcal{S} \neq \emptyset} P\left(\{U_1 \geq U_k \forall k\} \cap \{VU_j \geq VU_1 \text{ for at least one } j \in \mathcal{S}\}\right. \\
&\left.\cap \{g_1(x_1, U_j) \leq 0 \text{ for all } j \in \mathcal{S}\} \cap \{g_1(x_1, U_j) \geq 0 \text{ for all } j \in \mathcal{J}_1 \backslash \mathcal{S}\}\right) \\
&\equiv \ P_4(\mathbf{v}) - \sum_{\mathcal{S} \subset \mathcal{J}_1, \mathcal{S} \neq \emptyset} P_5^{\mathcal{S}}\left(\mathbf{v}, \mathbf{v}^0, x_1\right)
\end{aligned} \tag{20}$$

Further, for every $\mathcal{S} \subset \mathcal{J}_1, \mathcal{S} \neq \emptyset$, we have

$$\begin{aligned}
P_5^{\mathcal{S}} = \; & P\left(\{U_1 \geq U_k \; \forall k\} \cap \{g_1(x_1, U_j) \leq 0 \text{ for all } j \in \mathcal{S}\} \cap \{g_1(x_1, U_j) \geq 0 \text{ for all } j \in \mathcal{J}_1 \backslash \mathcal{S}\}\right) - \\
& P\left(\{U_1 \geq U_k \; \forall k\} \cap \{VU_1 \geq VU_j \text{ for all } j \in \mathcal{S}\}\right. \\
& \left. \cap \{g_1(x_1, U_j) \leq 0 \text{ for all } j \in \mathcal{S}\} \cap \{g_1(x_1, U_j) \geq 0 \text{ for all } j \in \mathcal{J}_1 \backslash \mathcal{S}\}\right) \\
= \; & P\left(\{U_1 \geq U_k \; \forall k\} \cap \{g_1(x_1, U_j) \leq 0 \text{ for all } j \in \mathcal{S}\} \cap \{g_1(x_1, U_j) \geq 0 \text{ for all } j \in \mathcal{J}_1 \backslash \mathcal{S}\}\right) - \\
& P\left(\{VU_1 \geq VU_j \text{ for all } j \in \mathcal{S}\} \cap \{g_1(x_1, U_j) \leq 0 \text{ for all } j \in \mathcal{S}\} \cap \{g_1(x_1, U_j) \geq 0 \text{ for all } j \in \mathcal{J}_1 \backslash \mathcal{S}\}\right) \\
\equiv \; & P_{5,1}^{\mathcal{S}}(\mathbf{v}, x_1) - P_{5,2}^{\mathcal{S}}\left(\mathbf{v}_{-1}, \mathbf{v}^0, x_1\right)
\end{aligned}$$

where $\mathbf{v}_{-1} \equiv (v_2, \ldots, v_J)$. 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) $VU_1 \geq VU_j$ implies $U_1 \geq U_j$ since $z_1 \geq z_j$ for all $j \in \mathcal{J}_1$; and (ii) $g_1(x_1, U_k) \geq 0 \geq g_1(x_1, U_j)$ 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.

Note that $P_{5,2}^{\mathcal{S}}$ does not depend on z_1. Thus, omitting the function arguments, we have

$$\frac{\partial s_1}{\partial z_1} = \frac{\partial P_4}{\partial v_1} \frac{\partial v_1}{\partial z_1} - \sum_{\mathcal{S} \subset \mathcal{J}_1, \mathcal{S} \neq \emptyset} \frac{\partial P_{5,1}^{\mathcal{S}}}{\partial v_1} \frac{\partial v_1}{\partial z_1} \tag{21}$$

Differentiating again with respect to z_2 gives:

$$\frac{\partial^2 s_1}{\partial z_1 \partial z_2} = \frac{\partial^2 P_4}{\partial v_1 \partial v_2} \frac{\partial v_1}{\partial z_1} \frac{\partial v_2}{\partial z_2} - \sum_{\mathcal{S} \subset \mathcal{J}_1, \mathcal{S} \neq \emptyset} \frac{\partial^2 P_{5,1}^{\mathcal{S}}}{\partial v_1 \partial v_2} \frac{\partial v_1}{\partial z_1} \frac{\partial v_2}{\partial z_2} \tag{22}$$

Differentiating equation (21) with respect to x_2 gives:

$$\frac{\partial^2 s_1}{\partial z_1 \partial x_2} = \frac{\partial^2 P_4}{\partial v_1 \partial v_2} \frac{\partial v_1}{\partial z_1} \frac{\partial v_2}{\partial x_2} - \sum_{\mathcal{S} \subset \mathcal{J}_1, \mathcal{S} \neq \emptyset} \frac{\partial^2 P_{5,1}^{\mathcal{S}}}{\partial v_1 \partial v_2} \frac{\partial v_1}{\partial z_1} \frac{\partial v_2}{\partial x_2} \tag{23}$$

Combining (22) and (23), we obtain

$$\frac{\partial^2 s_1}{\partial z_1 \partial z_2} \Big/ \frac{\partial^2 s_1}{\partial z_1 \partial x_2} = \frac{\frac{\partial v_2}{\partial z_2}}{\frac{\partial v_2}{\partial x_2}} \tag{24}$$

Since this equation holds for all $(\mathbf{x}, \mathbf{z})$ such that $\frac{\partial^2 s_1}{\partial z_1 \partial x_2} \neq 0$ and we already showed that we can recover $\frac{\partial v}{\partial x}(0,0)$, we can also recover $\frac{\partial v}{\partial z}(0,0)$.

Next, note that, fixing $z_k = 0$ for all $k = 1, \ldots, J$ and $x_j = 0$ for all $j \neq 2$ in (21), we can write

$$\frac{\partial s_1}{\partial z_1} = k(l(x_2)) \tag{25}$$

where

$$k(v_2): v_2 \mapsto \frac{\partial v}{\partial z_1}(\mathbf{0})\left[\frac{\partial P_4}{\partial v_1}(v(\mathbf{0}), v_2, v(\mathbf{0}), \ldots, v(\mathbf{0})) - \sum_{\mathcal{S}\subset\mathcal{J}_1, \mathcal{S}\neq\emptyset} \frac{\partial P_{5,1}^{\mathcal{S}}}{\partial v_1}(v(\mathbf{0}), v_2, v(\mathbf{0}), \ldots, v(\mathbf{0}), 0)\right]$$

and $l(x_2): x_2 \mapsto v(x_2, 0)$. So by the chain rule we have that, for $n > 1$, $\frac{\partial^n s_1}{\partial z_1 \partial x_2^{n-1}}$ is a linear function of the $(n-1)-$th derivative of k with slope depending on the first derivative of l and intercept depending on derivatives of l and derivatives of k of order strictly less than $n-1$. Further, by the above, all derivatives of l are known. Thus, we have a system of equations that can be uniquely solved for the derivatives of k by recursion.[40]

Next, we differentiate $\frac{\partial s_1}{\partial z_1}$ once with respect to z_2 and $n-2$ times with respect to x_2. Similar to the above, we can write

$$\frac{\partial s_1}{\partial z_1} = k(v(x_2, z_2)) \tag{26}$$

where now note that z_2 is no longer fixed at 0. Again by the chain rule we have that, for $n \geq 3$, $\frac{\partial^n s_1}{\partial z_1 \partial z_2 \partial x_2^{n-2}}$ evaluated at $(\mathbf{0}, \mathbf{0})$ is a linear function of $\frac{\partial^{n-1} v}{\partial z_2 \partial x_2^{n-2}}(0,0)$ with slope coefficient depending on $k'(v(0,0))$ and intercept depending on lower-order derivatives of v as well as derivatives of k.[41] Because all derivatives of k are known by the argument above, we can iteratively solve for $\frac{\partial^{n-1} v}{\partial z_2 \partial x_2^{n-2}}(0,0)$ for all $n \geq 3$.

The remaining terms in the Taylor expansion can be recovered by an analogous argument. Specifically, for any $n \geq 3, m \geq 2$, by differentiating (26) m times wrt z_2 and again $n-m-1$ times wrt x_2, one can write $\frac{\partial^n s_1}{\partial z_1 \partial z_2^m \partial x_2^{n-m-1}}$ as a linear function of $\frac{\partial^{n-1} v}{\partial z_2^m \partial x_2^{n-m-1}}(0,0)$ with known, nonzero slope and known intercept. This system can then be solved iteratively for $\frac{\partial^{n-1} v}{\partial z_2^m \partial x_2^{n-m-1}}(0,0)$ for all $n > m \geq 2$.

Therefore, we know all the coefficients in the Taylor-expansion of $v(x,z)$ except the constant $v(0,0)$, i.e. we can recover $v(x,z)$ up to a constant.

A.2 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 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 z attributes. This is not without loss.

We now show how to relax this restriction in the linear homogeneous case of Lemma 3. Let z_{kj} be the $k-$th hidden attribute of good j and let β_k be the associated preference parameter. By

[40] Here, we use the fact that, by assumption, the first derivative of l is nonzero.

[41] Note that $\frac{\partial^2 s_1}{\partial z_1 \partial x_2}(\mathbf{0}, \mathbf{0}) = k'(v(0,0))\, l'(0)$, so we have $k'(v(0,0)) \neq 0$ by assumption.

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', $\sum_k \beta_k z_{kj} \geq \sum_k \beta_k z_{kj'}$ if and only if $z_{1j} - z_{1j'} + \sum_{k>1} \frac{\beta_k}{\beta_1}(z_{kj} - z_{kj'}) \geq 0$. Since the l.h.s. of the last inequality is identified, we can rank goods based on their non-visible utility. Lemma 3 then applies by defining good 1 as the good with the highest value of $\sum_k \beta_k z_{kj}$. Note that such a good always exists in any choice set (excluding ties) since $\sum_k \beta_k z_{kj}$ is scalar-valued.

A.3 Proof of Theorem 3

Note that the analog of equation (20) in the random coefficients model holds for any given value of α and $\beta > 0$, so that we can write:

$$s_1 = \sum_{k_\alpha=1}^{K_\alpha} \sum_{k_\beta=1}^{K_\beta} \left[P_6(\mathbf{v}(k_\alpha, k_\beta)) - \sum_{\mathcal{S} \subset \mathcal{J}_1, \mathcal{S} \neq \emptyset} P_7^{\mathcal{S}}\left(\mathbf{v}(k_\alpha, k_\beta), \mathbf{v}^0(k_\alpha), x_1\right) \right] \tilde{\pi}_{k_\alpha, k_\beta} \tag{27}$$

where $v_j(k_\alpha, k_\beta) \equiv x_j \alpha_{k_\alpha} + z_j \beta_{k_\beta}$, $\mathbf{v}(k_\alpha, k_\beta) \equiv (v_1(k_\alpha, k_\beta), \ldots, v_J(k_\alpha, k_\beta))$, and similarly $v_j^0(k_\alpha) \equiv x_j \alpha_{k_\alpha}$, $\mathbf{v}^0(k_\alpha) \equiv \left(v_1^0(k_\alpha), \ldots, v_J^0(k_\alpha)\right)$. Differentiating (27), we have, for all integers $n \geq \tilde{n} \geq 0$:

$$\frac{\partial^{1+n} s_1}{\partial z_1 \partial z_2^{\tilde{n}} \partial x_2^{n-\tilde{n}}} = \sum_{k_\alpha=1}^{K_\alpha} \sum_{k_\beta=1}^{K_\beta} \frac{\partial^{1+n} \left[P_6(\mathbf{v}(k_\alpha, k_\beta)) - \sum_{\mathcal{S} \subset \mathcal{J}_1, \mathcal{S} \neq \emptyset} P_7^{\mathcal{S}}\left(\mathbf{v}(k_\alpha, k_\beta), \mathbf{v}^0(k_\alpha), x_1; \alpha_{k_\alpha}, \beta_{k_\beta}\right) \right]}{\partial z_1 \partial z_2^{\tilde{n}} \partial x_2^{n-\tilde{n}}} \tilde{\pi}_{k_\alpha, k_\beta} \tag{28}$$

Next, we evaluate (28) at a value of $(\mathbf{x}, \mathbf{z})$ such that $x_j = \bar{x}$ and $z_j = \bar{z}_j$ for all j. Note that at such $(\mathbf{x}, \mathbf{z})$, P_6 no longer depends on k_α, k_β. Using this and letting $\bar{x} = \bar{z} = 0$ for notational convenience (the proof is unchanged for any other values), we may re-write (28) as

$$\begin{aligned} \frac{\partial^{1+n} s_1}{\partial z_1 \partial z_2^{\tilde{n}} \partial x_2^{n-\tilde{n}}} &= \frac{\partial^{1+n} P_6}{\partial v_1 \partial v_2^n}(\mathbf{0}) \sum_{k_\alpha=1}^{K_\alpha} \sum_{k_\beta=1}^{K_\beta} \alpha_{k_\alpha}^{n-\tilde{n}} \beta_{k_\beta}^{\tilde{n}+1} \tilde{\pi}_{k_\alpha, k_\beta} - \sum_{k_\alpha=1}^{K_\alpha} \sum_{k_\beta=1}^{K_\beta} \alpha_{k_\alpha}^{n-\tilde{n}} \beta_{k_\beta}^{\tilde{n}+1} \tilde{\pi}_{k_\alpha, k_\beta} \sum_{\mathcal{S} \subset \mathcal{J}_1, \mathcal{S} \neq \emptyset} \frac{\partial^{1+n} P_7^{\mathcal{S}}\left(\mathbf{0}, \mathbf{0}, 0; \alpha_{k_\alpha}, \beta_{k_\beta}\right)}{\partial v_1 \partial v_2^n} \\ &\equiv \sum_{k=1}^{K} \left[a_{k,n,\tilde{n}} + b_{k,n,\tilde{n}} f_{k,n}\right] \pi_k \end{aligned} \tag{29}$$

where we let $K \equiv K_\alpha K_\beta$ and let k represent the double index (k_α, k_β), $a_{k,n,\tilde{n}} \equiv \frac{\partial^{1+n} P_6}{\partial v_1 \partial v_2^n}(\mathbf{0}) \alpha_{k_\alpha}^{n-\tilde{n}} \beta_{k_\beta}^{\tilde{n}+1}$, $b_{k,n,\tilde{n}} \equiv \alpha_{k_\alpha}^{n-\tilde{n}} \beta_{k_\beta}^{\tilde{n}+1}$ are known scalars, and $f_{k,n} \equiv \frac{\partial^{1+n} P_7^{\mathcal{S}}\left(\mathbf{0}, \mathbf{0}, 0; \alpha_{k_\alpha}, \beta_{k_\beta}\right)}{\partial v_1 \partial v_2^n}$ and $\pi_k \equiv \tilde{\pi}_{k_\alpha, k_\beta}$ are unknown scalars.

Setting $n = K - 1$ and stacking the equations corresponding to $\tilde{n} = 0, \ldots, K-1$, we get

$$q = A\pi + B(f * \pi)$$

where q is a known column $K-$vector, A, B are known $K-$by$-K$ matrices, and $f * \pi$ denotes the column vector given by the element-by-element product of $f = (f_{1,K-1}, \ldots, f_{K,K-1})'$ and $\pi \equiv (\pi_1, \ldots, \pi_K)'$.

We re-write this system of equations in a way that highlights which objects depend on $\mathbf{z} \equiv (z_1, \ldots, z_J)$ as follows

$$q(\mathbf{z}=\mathbf{0}) \quad = \quad A\pi + B(f(\mathbf{z}=\mathbf{0}) * \pi)$$

Note that A depends on z only through $z_1 - z_j$ (i.e. it exhibits a lack of nominal illusion property) and we leave that dependence implicit. Now consider increasing z_j by Δz for all j relative to the baseline $\mathbf{z}=\mathbf{0}$. Then we can write

$$q(\mathbf{z}=\mathbf{\Delta z}) \quad = \quad A\pi + B(f(\mathbf{z}=\mathbf{\Delta z}) * \pi)$$

Combining the last two systems, we get

$$q(\mathbf{z}=\mathbf{\Delta z}) - q(\mathbf{z}=\mathbf{0}) \quad = \quad B\left[(f(\mathbf{z}=\mathbf{\Delta z}) - f(\mathbf{z}=\mathbf{0})) * \pi\right]$$

If B is full rank,[42] we obtain identification of $(f(\mathbf{z}=\mathbf{\Delta z}) - f(\mathbf{z}=\mathbf{0})) * \pi$. Also, note that, for all k, $\lim_{\Delta z \to 0} \frac{f_{k,K-1}(\mathbf{z}=\mathbf{\Delta z}) - f_{k,K-1}(\mathbf{z}=\mathbf{0})}{\Delta z}$ is the directional derivative of $f_{k,K-1}$ in the direction $\mathbf{1} = (1, \ldots, 1)$ and thus is equal to $\sum_{j=1}^{J} \frac{\partial f_{k,K-1}}{\partial z_j}(\mathbf{z}=\mathbf{0})$ if $f_{k,K-1}$ is differentiable. Therefore, we can write

$$\lim_{\Delta z \to 0} \frac{q(\mathbf{z}=\mathbf{\Delta z}) - q(\mathbf{z}=\mathbf{0})}{\Delta z} \quad = \quad B\left[\left(\sum_{j=1}^{J} \frac{\partial f}{\partial z_j}(\mathbf{z}=\mathbf{0})\right) * \pi\right]$$

Because the lhs is identified, this shows that we can identify $\left(\sum_{j=1}^{J} \frac{\partial f}{\partial z_j}(\mathbf{z}=\mathbf{0})\right) * \pi$.

Next, for $j \in \mathcal{J}$, we can take another derivative wrt z_j in (29) and write

$$q_{(j)}(\mathbf{z}=\mathbf{0}) = A_{(j)}\pi + B_{(j)}\left(\frac{\partial f}{\partial z_j}(\mathbf{z}=\mathbf{0}) * \pi\right) \tag{30}$$

for known K−by−K matrices $A_{(j)}, B_{(j)}$ and a known column K−vector $q_{(j)}(\mathbf{z}=\mathbf{0})$. Note that $B_{(j)} = B$ for all $j \in \mathcal{J}$ and so we can write

$$\sum_{j=1}^{J} q_{(j)}(\mathbf{z}=\mathbf{0}) = \left(\sum_{j=1}^{J} A_{(j)}\right)\pi + B\left[\left(\sum_{j=1}^{J} \frac{\partial f}{\partial z_j}(\mathbf{z}=\mathbf{0})\right) * \pi\right] \tag{31}$$

From above, $\left(\sum_{j=1}^{J} \frac{\partial f}{\partial z_j}(\mathbf{z}=\mathbf{0})\right) * \pi$ is identified. This implies that π is identified if the matrix $\sum_{j=1}^{J} A_{(j)}$ is invertible.[43]

[42] Note that this condition is immediately verifiable since the points in the support of α and β are chosen by the researcher.

[43] Again, this condition is immediately verifiable given the support points for α and β chosen by the researcher.

A.4 Endogenous attributes

Here, we show how to extend our results to the case where some product attributes are endogenous (Section 3.1). Letting $\delta = (\delta_1, \cdots, \delta_J)$, we may write the share of good j as

$$s_j = \sigma_j (\delta, \mathbf{z}, \mathbf{p}) \tag{32}$$

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

$$\mathbf{s} = \sigma (\delta, \mathbf{z}, \mathbf{p}) \tag{33}$$

where $\mathbf{s} = (s_1, \cdots, s_J)$. We also define the share of the outside option as $s_0 \equiv 1 - \sum_{j=1}^{J} s_j$, with associated function $\sigma_0 (\delta, \mathbf{z}, \mathbf{p})$. We establish nonparametric identification of this demand system by invoking results from Berry and Haile (2014) (henceforth, BH).[44] Specifically, the results in BH yield identification of $(\xi_j)_{j=1}^{J}$ for every unit (individual or 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 2.3 to identify the distribution of the preference parameters α, β_i and λ_i. 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 (32) 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

$$E (\xi_j | \mathbf{x}, \mathbf{z}, \mathbf{w}) = 0 \quad \text{for all } j \tag{34}$$

almost surely (Assumption 3 in BH) and assume that the instruments shift the endogenous variables (market shares and endogenous attributes $\mathbf{p}$) to a sufficient degree (as in BH's Assumption 4). 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 4. *Let utility be given by* (13) *with ϵ_i supported on $\mathbb{R}^J$ and let Assumptions 2(i), 2(iii), 2(iv), and either 4(i) or 4(ii) hold. Then, for all $j, k = 1, \cdots, J$ with $j \neq k$, σ_j is (i) strictly increasing in δ_j and (ii) strictly decreasing in δ_k.*

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

Proof. First, assume that p_j is part of the visible utility of good j and Fix (δ_j, p_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 $g_{ij}(\delta_j, p_j, U_{ik}) \geq 0$ for all k such that $VU_{ik} \geq VU_{ij}$. Let $\Delta \geq 0$ be the change in δ_j. Since g_{ij} is increasing in its first argument, we have $g_{ij}(\delta_j + \Delta, p_j, U_{ik}) \geq 0$ for all k such that $VU_{ik} \geq VU_{ij} + \Delta$ and thus i will still search j. Moreover, since g_{ij} is decreasing in its last argument, if $g_{ik}(\delta_k, p_k, U_{ij}) \leq 0$ for some k such that $VU_{ik} \leq VU_{ij}$ (i.e. if k is initially not searched), then $g_{ik}(\delta_k, p_k, U_{ij} + \Delta) \leq 0$ (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).

Since the argument above does not rely on the fact that p_j is part of the visible utility of good j, the conclusion also holds for the case in which p_j is only uncovered upon searching good j. □

Lemma 4 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.[45] Since Lemma 4 holds under either Assumption 4(i) or 4(ii), we obtain identification of preferences both in the case where p_j is part of the visible utility of good j and in the case where p_j is only uncovered upon searching j. Moreover, Theorem 1 of BH implies that one can invert the demand system σ for the indices δ and write

$$\alpha x_j + \xi_j = \sigma_j^{-1}(\mathbf{s}, \mathbf{z}, \mathbf{p}) \tag{35}$$

for all j. Equations (35) and (34) naturally lead to a nonparametric instrumental variable approach to estimate σ_j^{-1} (and thus σ_j).[46]

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 5. (i) If consumer i searches j, then i also searches all j' s.t. $m\left(VU_{ij'}, r_{j'}\right) \geq m\left(VU_{ij}, r_j\right)$, 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 $(\mathbf{x}, \mathbf{z}) \Big| (r_1, \ldots, r_J)$ has positive Lebesgue measure for all $(r_1, \ldots, r_J)$.

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

[46] Compiani (2019) proposes to approximate σ_j^{-1} using Bernstein polynomials. We use a similar approach in Section 4 to estimate the demand function for the case without endogeneity.

Assumption 5(*iii*) is substantive: 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.

Violations of the visible utility assumption due to search position will cause Lemma 1 to no longer hold as stated: the good with the highest value of z_j can be searched, another good j' may have higher utility (and thus higher visible 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 5. *Let Assumptions 1, 2(ii)-2(iv), and 5 hold and let $x_j \in [\bar{x} - \eta, \bar{x} + \eta]$ for all j, for some $\eta > 0$ sufficiently small. Then, if consumer i searches good 1 (i.e. the good with the highest value of 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 $r_j \geq r_1$ and $U_{ij} > U_{i1}$ that consumer i does not search. We can follow the proof of Lemma 1 to show that $VU_{ij} > VU_{i1}$. By Assumption 5(i), this implies that good j is searched, which is a contradiction. □

In other words, if higher search position only makes a good more likely to be searched, then goods with higher visible 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 Theorem 1 after conditioning on the subset of goods with higher search position than good 1 (defined as usual as the good with the largest value of z_j):

Theorem 4. *Let the assumptions of Lemma 5 hold and let utility be given by $U_{ij} = v(x_j, z_j) + \epsilon_{ij}$ with v increasing in both arguments and infinitely differentiable. Further, assume 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 $j^* \neq 1$, s_1 is infinitely differentiable and $\epsilon_i \perp (\mathbf{x}, \mathbf{z})$. Then, v is identified up to an additive constant.*

Proof. Note that if good 1 doesn't maximize utility in R, good 1 will never be chosen. If some other good in R has higher utility, it has higher visible utility and will be searched before good 1. Then, by Lemma 5, we can write for all $(\mathbf{x}, \mathbf{z})$ with $z_1 \geq z_j$ and all j:

$$\begin{aligned} s_1 = {} & P(U_1 \geq U_k \ \forall k \in R) - \sum_{\mathcal{S} \subset \mathcal{J}_1, \mathcal{S} \neq \emptyset} P\left(\{U_1 \geq U_k \forall k \in R\} \cap \{m(VU_j, r_j) \geq m(VU_1, r_1) \text{ for at least one } j \in \mathcal{S}\}\right. \\ & \left. \cap \{g_1(x_1, U_j) \leq 0 \text{ for all } j \in \mathcal{S}\} \cap \{g_1(x_1, U_j) \geq 0 \text{ for all } j \in \mathcal{J}_1 \backslash \mathcal{S}\}\right) \\ \equiv {} & P_{4new}(\mathbf{v}) - \sum_{\mathcal{S} \subset \mathcal{J}_1, \mathcal{S} \neq \emptyset} P_{5new}^{\mathcal{S}}\left(\mathbf{v}, \mathbf{v}^0, x_1, \mathbf{r}\right) \end{aligned} \tag{36}$$

where $\mathbf{v}$ and $\mathbf{v}^0$ are defined as in the proof of Theorem 1 (Appendix A.1) and $\mathbf{r} \equiv (r_1, \ldots, r_J)$. Further, for every $\mathcal{S} \subset \mathcal{J}_1, \mathcal{S} \neq \emptyset$, we have

$$
\begin{aligned}
P^{\mathcal{S}}_{5new} = \; & P\left(\{U_1 \geq U_k \;\forall k \in R\} \cap \{g_1(x_1, U_j) \leq 0 \text{ for all } j \in \mathcal{S}\} \cap \{g_1(x_1, U_j) \geq 0 \text{ for all } j \in \mathcal{J}_1 \backslash \mathcal{S}\}\right) - \\
& P\left(\{U_1 \geq U_k \;\forall k \in R\} \cap \{m(VU_1, r_1) \geq m(VU_j, r_j) \text{ for all } j \in \mathcal{S}\}\right. \\
& \left. \cap \{g_1(x_1, U_j) \leq 0 \text{ for all } j \in \mathcal{S}\} \cap \{g_1(x_1, U_j) \geq 0 \text{ for all } j \in \mathcal{J}_1 \backslash \mathcal{S}\}\right) \\
= \; & P\left(\{U_1 \geq U_k \;\forall k \in R\} \cap \{g_1(x_1, U_j) \leq 0 \text{ for all } j \in \mathcal{S}\} \cap \{g_1(x_1, U_j) \geq 0 \text{ for all } j \in \mathcal{J}_1 \backslash \mathcal{S}\}\right) - \\
& P\left(\{m(VU_1, r_1) \geq m(VU_j, r_j) \text{ for all } j \in \mathcal{S}\} \cap \{g_1(x_1, U_j) \leq 0 \text{ for all } j \in \mathcal{S}\} \cap \{g_1(x_1, U_j) \geq 0 \text{ for all } j \in \mathcal{J}_1 \backslash \mathcal{S}\}\right) \\
\equiv \; & P^{\mathcal{S}}_{5new,1}(\mathbf{v}, x_1, \mathbf{r}) - P^{\mathcal{S}}_{5new,2}(\mathbf{v}_{-1}, \mathbf{v}^0, x_1, \mathbf{r})
\end{aligned}
$$

where $\mathbf{v}_{-1} \equiv (v_2, \ldots, v_J)$. This argument exactly parallels the argument in Appendix A.1, except now we have additionally used the fact that $U_1 \geq U_j$ for all $j \in R$, since (i) if $j \in \mathcal{S}$, then $m(VU_1, r_1) \geq m(VU_j, r_j)$ implies $VU_1 \geq VU_j$, which in turn implies $U_1 \geq U_j$; (ii) if $j \notin \mathcal{S}$, then $g_1(x_1, U_j) \geq 0 \geq g_1(x_1, U_k)$ for all $k \in \mathcal{S}$ implies $U_j \leq U_k$. Note that $P^{\mathcal{S}}_{5new,2}$ does not depend on z_1 and $P^{\mathcal{S}}_{5new,1}(\mathbf{v}, x_1, \mathbf{r})$ only depends on x_j and z_j via v_j for $j \neq 1$, so the remainder of the argument in Appendix A.1 applies.

□

A.6 Identification of a model where consumers form expectations on z_j based on x_j

Here, we state and prove the results described in Section 3.3. Given γ_1, we can identify the ranking of goods in terms of $\tilde{z}$ and we label good 1 as the good with the largest value of $\tilde{z}$. Then, an argument analogous to that in Lemma 3 yields identification of $\frac{\beta}{\alpha+\beta\gamma_1}$.[47] We can also recover $\alpha+\beta\gamma_1$ in a manner that parallels our usual identification of α (Lemma 3). When $\tilde{z}_j = \tilde{z}$ for all j, consumers who search based on our visible utility assumption always maximize utility, and thus we can directly estimate $\alpha + \beta\gamma_1$ as the coefficient on x_j for those consumers (we provide a formal proof of this in the next subsection). Therefore, this gives separate identification of β and α given γ_1.

When γ_1 is unknown, we can identify β/α if we know its sign and make a further support assumption. Suppose that the sign of γ_1 is known (e.g. higher priced goods have weakly 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 $\tilde{z}$; thus, we can label it by 1. Note that we cannot differentiate with respect to $\tilde{z}$ as in the case above since γ_1 and thus $\tilde{z}$ is unknown. However, with good 1 defined appropriately, Corollary 1 shows that cross-derivatives with respect to z_1, z_j, x_j for $j \neq 1$ identify β/α (specifically, consumers who search the good with the highest value of $\tilde{z}$ will always maximize utility, and so their sensitivity to x_j and z_j identifies their true preferences).

[47] In Lemma 3, we showed identification of $\frac{\beta}{\alpha}$ by taking derivatives of s_1 w.r.t. z_1, z_2, x_2. Similarly, here we obtain identification of $\frac{\beta}{\alpha+\beta\gamma_1}$ by taking derivatives of s_1 wrt $\tilde{z}_1, \tilde{z}_2, x_2$.

A.6.1 Identification of $\alpha + \beta\gamma_1$

Note that if $\tilde{z}_j = 0$ for all j, then consumers always maximize utility. Thus, seeing how choice probabilities change with x conditional on $\tilde{z}_j = 0$ for all j should help identify $\alpha + \beta\gamma_1$. Because the event $\tilde{z}_j = 0$ involves x_j, we need to differentiate choice probabilities with respect to x_j on the envelope satisfying the condition $\tilde{z}_j = 0$ for all x_j. Formally, fix any $j \in \mathcal{J}$ and choose (x_k, z_k) so that $z_k = \gamma_0 + \gamma_1 x_k$ (which implies $\tilde{z}_k = 0$) for all $k \neq j$. For every $\delta > 0$, let $\epsilon(\delta) \equiv \gamma_0 + (x_j + \delta)\gamma_1 - z_j$, so that $z_j + \epsilon(\delta) - E(z_j | x_j + \delta) = 0$. Note that $\epsilon(\delta)$ is known to the econometrician. Denoting by $\mathbf{x}_{-j} = (x_k)_{k \neq j}$ and similarly for $\mathbf{z}_{-j}$, we have

$$
\begin{aligned}
& \frac{s_j (x_j + \delta, \mathbf{x}_{-j}, z_j + \epsilon(\delta), \mathbf{z}_{-j}) - s(\mathbf{x}, \mathbf{z})}{\delta} && (37) \\
= \quad & \frac{P((x_j + \delta)(\alpha + \beta\gamma_1) + \epsilon_{ij} \geq x_k (\alpha + \beta\gamma_1) + \epsilon_{ik} \forall k) - P(x_j (\alpha + \beta\gamma_1) + \epsilon_{ij} \geq x_k (\alpha + \beta\gamma_1) + \epsilon_{ik} \forall k)}{\delta} \\
\xrightarrow{\delta \to 0} \quad & \frac{\partial}{\partial x_j} P(x_j (\alpha + \beta\gamma_1) + \epsilon_{ij} \geq x_k (\alpha + \beta\gamma_1) + \epsilon_{ik} \ \forall k) && (38)
\end{aligned}
$$

where the first equality follows from the fact that all consumers always maximize utility at the chosen values of $(\mathbf{x}, \mathbf{z})$. Now note that (i) the expression in (37) is known for all $\delta > 0$; and (ii) at $\mathbf{x} = \mathbf{0}$, the term in (38) factors into $(\alpha + \beta\gamma_1)$ and a term that only depends the distribution of ϵ_i. Thus, evaluating the last display at $\mathbf{x} = \mathbf{0}$ yields identification of $(\alpha + \beta\gamma_1)$ under a parametric assumption on ϵ_i.

A.7 Unobservables revealed by search

Here, we show that the ratio of second derivatives in (5) robustly identifies $\frac{\beta}{\alpha}$ in the model where ϵ_{ij} is revealed to consumer i only upon searching good j (Section 3.4).

Order goods in increasing order of x. Then, for $j = 1, \ldots, J$,

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

Thus,

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

So 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). On the other hand,

$$\begin{aligned}\frac{\partial s_j}{\partial z_j} &= \sum_{k=1}^{j} \frac{\partial P_j^{(k)}}{\partial \tilde{u}_j}\beta \\ \frac{\partial s_j}{\partial x_j} &= \sum_{k=1}^{j} \left(\frac{\partial P_j^{(k)}}{\partial \tilde{u}_j} + \frac{1}{\alpha}\frac{\partial P_j^{(k)}}{\partial x_j} \right)\alpha\end{aligned} \tag{39}$$

Since $\frac{1}{\alpha}\frac{\partial P_j^{(k)}}{\partial x_j} \geq 0$, (39) 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.8 $K-$rank model

Consider the simultaneous search model in Honka, Hortaçsu, and Vitorino (2017) with $J = 2$ goods. In this model, a consumer looks at the visible utilities and decides whether to search the good with the highest visible utility or search both goods. Searching a second good entails a cost c, constant across consumers. As usual, we denote by 1 the good with the highest value of z.

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

$$E_{z_1,z_2}\left[\max\left\{VU_{i1} + \beta z_1, VU_{i2} + \beta z_2\right\}\right] - c < E_{z_2}\left[VU_{i2} + \beta z_2\right] \tag{40}$$

i.e.

$$E_{z_1,z_2}\left[\max\left\{VU_{i1} - VU_{i2} + \beta\left(z_1 - z_2\right), 0\right\}\right] - c < 0 \tag{41}$$

or $g_{sim}\left(VU_{i1} - VU_{i2}\right) < 0$ for an increasing function g_{sim}. Equation (6) then can be written as

$$\begin{aligned}s_1 &= P(U_1 > U_2) - P(\{U_1 > U_2\} \cap \{VU_2 > VU_1\} \cap \{g_{sim}(VU_1 - VU_2) < 0\}) \\ &= P_{1,sim} - P_{2,sim}\end{aligned} \tag{42}$$

We also have:

$$\frac{\partial^2 s_1}{\partial z_1 \partial z_2} = \beta^2 \left(\frac{\partial^2 P_{1,sim}}{\partial \tilde{u}_1 \partial \tilde{u}_2} - \frac{\partial^2 P_{2,sim}}{\partial \tilde{u}_1 \partial \tilde{u}_2} \right) \tag{43}$$

and

$$\frac{\partial^2 s_1}{\partial z_1 \partial x_2} = \alpha\beta \left(\frac{\partial^2 P_{1,sim}}{\partial \tilde{u}_1 \partial \tilde{u}_2} - \frac{\partial^2 P_{2,sim}}{\partial \tilde{u}_1 \partial \tilde{u}_2} - \frac{\partial^2 P_{2,sim}}{\partial \tilde{u}_1 \partial \tilde{vu}_2} \right) \tag{44}$$

So, if $\frac{\partial^2 P_{2,sim}}{\partial \tilde{u}_1 \partial \tilde{vu}_2} = 0$, then the ratio of (43) to (44) 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) $\epsilon_{i1} < \tilde{vu}_2 - \tilde{vu}_1 + \epsilon_{i2}$, (iii) $\epsilon_{i1} < g_{sim}^{-1}(0) + \tilde{vu}_2 - \tilde{vu}_1 + \epsilon_{i2}$, where $VU_{ij} = \tilde{vu}_j + \epsilon_{ij}$ and $U_{ij} = \tilde{u}_j + \epsilon_{ij}$, as above. Then, letting $\tilde{\epsilon} = \epsilon_1 - \epsilon_2$, we have:

$$P_{2,sim} = \int_{\tilde{u}_2 - \tilde{u}_1}^{\min(\tilde{vu}_2 - \tilde{vu}_1, g_{sim}^{-1}(0) + \tilde{vu}_2 - \tilde{vu}_1)} f_{\tilde{\epsilon}}(\tilde{\epsilon}) d\tilde{\epsilon} = \int_{\beta(z_2 - z_1)}^{\min(0, g_{sim}^{-1}(0))} f_{\tilde{\epsilon}}(\tilde{\epsilon}) d\tilde{\epsilon}$$

Thus, $\frac{\partial^2 P_{2,sim}}{\partial \tilde{u}_1 \partial \tilde{vu}_2} = 0$.

Finally, we show that the ratio of first derivatives leads to attenuation bias. This follows directly from

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

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

Supplemental Appendices

Appendix B: Testing for full information with heterogeneous preferences

In Section 2.7, 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, a valid test rejects the null when the ratios of first derivatives are attenuated relative to the ratio of second derivatives in (12). Here, we provide conditions under which the same test is valid in the case where one of the two coefficients is allowed to be heterogeneous.[48] We focus on the case where β is heterogeneous and z_j is a scalar; the argument for the case where α is heterogeneous (and x_j is a scalar) is analogous. We also assume that the ϵ_{ij} shocks are type-I extreme-value distributed and let $s_j(\tilde{\beta})$ be the market share of good 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 let $j = 2, k = k' = 1$ in equation (12), i.e. we consider the case where the test compares the ratio of second derivatives taken with respect to good 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 j, k, k'. Then, we want to show that

$$\frac{\int s_1(\mathbf{x}, \mathbf{z}; \beta)(1 - s_1(\mathbf{x}, \mathbf{z}; \beta))\beta dF_\beta}{\alpha \int s_1(\mathbf{x}, \mathbf{z}; \beta)(1 - s_1(\mathbf{x}, \mathbf{z}; \beta))dF_\beta} \geq \frac{-\int s_1(\mathbf{x}, \mathbf{z}; \beta)s_2(\mathbf{x}, \mathbf{z}; \beta)(1 - 2s_1(\mathbf{x}, \mathbf{z}; \beta))\beta^2 dF_\beta}{-\alpha \int s_1(\mathbf{x}, \mathbf{z}; \beta)s_2(\mathbf{x}, \mathbf{z}; \beta)(1 - 2s_1(\mathbf{x}, \mathbf{z}; \beta))\beta dF_\beta} \tag{45}$$

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 (45) holds if and only if

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

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

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

[48] 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.

if and only if, for all β,

$$\begin{aligned}
&(1-2s_1(\mathbf{x},\mathbf{z};\beta))(1-s_1(\mathbf{x},\mathbf{z};\beta)) > \\
\beta&\left[s_1(\mathbf{x},\mathbf{z};\beta)\left(z_1-\sum_{k=1}^{J}s_k(\mathbf{x},\mathbf{z};\beta)z_k\right)-(1-s_1(\mathbf{x},\mathbf{z};\beta))(1-2s_1(\mathbf{x},\mathbf{z};\beta))\left(z_2-\sum_{k=1}^{J}s_k(\mathbf{x},\mathbf{z};\beta)z_k\right)\right]
\end{aligned} \tag{46}$$

Under these conditions, at the chosen values of $\mathbf{x},\mathbf{z}$, a valid test of the null of full information rejects when the ratio of first derivatives is sufficiently attenuated relative to the ratio of first derivatives. Note that the condition in (46) can be verified given the support of the distribution of β. For example, if β takes values on a finite grid of points (as in Section 2.3), then one needs to check whether (46) holds for all values in the grid. Finally, we emphasize that (46) is a sufficient, but in general not necessary condition, implying that the proposed test could be valid even if the restriction is not satisfied.

Appendix C: Derivation of Flexible Logit Weights and Choice Probabilities

In this section, we derive the relevant derivatives of choice probabilities for the flexible logit model described in the text. In this model:

$$v_1 = \tilde{v}(x_1, z_1) + b_1 z_1 + \sum_{k\neq 1}(\gamma_k w_{z1k} z_k + \gamma_{2k} w_{x1k} x_k + w_{z2k}\delta_k z_k z_1 + w_{x2k}\delta_{2k} x_k z_1) \tag{47}$$

and $v_k = \tilde{v}(x_k, z_k)$ for $k \neq 1$ where b_1, γ_k, γ_{2k}, δ_k and δ_{2k} are coefficients to be estimated which allow greater flexibility in how derivatives with respect to z_1 vary with attributes of rival goods. 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 (where we use the notation $\tilde{v}_j$ to refer to the function $\tilde{v}$ evaluated at (x_j, z_j):

$$\begin{aligned}
\frac{\partial v_1}{\partial z_1} &= \frac{\partial \tilde{v}_1}{\partial z} + b_1 + \sum_{k\neq 1}(w_{z2k}\delta_k z_k + w_{x2k}\delta_{2k}x_k) \\
\frac{\partial s_1}{\partial x_1} &= \frac{\partial s_1}{\partial v_1}\frac{\partial v_1}{\partial x_1} = \frac{\partial \tilde{v}_1}{\partial x}s_1(1-s_1) \\
\frac{\partial s_1}{\partial z_1} &= \frac{\partial s_1}{\partial v_1}\frac{\partial v_1}{\partial z_1} = \frac{\partial v_1}{\partial z_1}s_1(1-s_1) \\
\frac{\partial s_1}{\partial x_{j'}} &= \frac{\partial s_1}{\partial v_{ij'}}\frac{\partial v_{ij'}}{\partial x_{j'}} + \frac{\partial s_1}{\partial v_{i1}}\frac{\partial v_{i1}}{\partial x_{j'}} = -\frac{\partial \tilde{v}_{j'}}{\partial x}s_1 s_{j'} + [w_{x1j'}\gamma_{2j'} + w_{x2j'}\delta_{2j'}z_1]s_1(1-s_1) \\
\frac{\partial s_1}{\partial z_{j'}} &= \frac{\partial s_1}{\partial v_{ij'}}\frac{\partial v_{ij'}}{\partial z_{j'}} + \frac{\partial s_1}{\partial v_{i1}}\frac{\partial v_{i1}}{\partial z_{j'}} = -\frac{\partial \tilde{v}_{j'}}{\partial z}s_1 s_{j'} + [w_{z1j'}\gamma_{j'} + w_{z2j'}\delta_{j'}z_1]s_1(1-s_1) \\
\\
\frac{\partial^2 s_1}{\partial z_1 \partial x_{j'}} &= \frac{\partial^2 s_1}{\partial v_1 \partial x_{j'}}\frac{\partial v_1}{\partial z_1} + \frac{\partial s_1}{\partial v_1}\frac{\partial^2 v_1}{\partial z_1 \partial x_{j'}} \\
&= \frac{\partial v_1}{\partial z_1}(1-2s_1)\frac{\partial s_1}{\partial x_{j'}} + s_1(1-s_1)w_{x2j'}\delta_{2j'} \\
\frac{\partial^2 s_1}{\partial z_1 \partial z_{j'}} &= \frac{\partial^2 s_1}{\partial v_1 \partial z_{j'}}\frac{\partial v_1}{\partial z_1} + \frac{\partial s_1}{\partial v_1}\frac{\partial^2 v_1}{\partial z_1 \partial z_{j'}} \\
&= \frac{\partial v_1}{\partial z_1}(1-2s_1)\frac{\partial s_1}{\partial z_{j'}} + s_1(1-s_1)w_{z2j'}\delta_{j'}
\end{aligned} \tag{48}$$

And also:

$$\begin{aligned}
\frac{\partial^2 s_1}{\partial z_1 \partial z_{j'}} / \frac{\partial^2 s_1}{\partial z_1 \partial x_{j'}} &= \frac{\frac{\partial v_1}{\partial z_1}(1-2s_1)\frac{\partial s_1}{\partial z_{j'}} + s_1(1-s_1)w_{z2j'}\delta_{j'}}{\frac{\partial v_1}{\partial z_1}(1-2s_1)\frac{\partial s_1}{\partial x_{j'}} + s_1(1-s_1)w_{x2j'}\delta_{2j'}} \\
&= \frac{\frac{\partial v_1}{\partial z_1}(1-2s_1)\left(-\frac{\partial \tilde{v}_{j'}}{\partial z}s_1 s_{j'} + [w_{z1j'}\gamma_{j'} + w_{z2j'}\delta_{j'}z_1]s_1(1-s_1)\right) + s_1(1-s_1)w_{z2j'}\delta_{j'}}{\frac{\partial v_1}{\partial z_1}(1-2s_1)\left(-\frac{\partial \tilde{v}_{j'}}{\partial x}s_1 s_{j'} + [w_{x1j'}\gamma_{2j'} + w_{x2j'}\delta_{2j'}z_1]s_1(1-s_1)\right) + s_1(1-s_1)w_{x2j'}\delta_{2j'}}
\end{aligned} \tag{49}$$

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

$$\begin{aligned}
\frac{\partial^2 s_1}{\partial z_1 \partial z_{j'}} / \frac{\partial^2 s_1}{\partial z_1 \partial x_{j'}} &= \frac{\frac{\partial v_1}{\partial z_1}(1-2s_1)s_1 s_{j'}\left(-\frac{\partial \tilde{v}_{j'}}{\partial z} + \gamma_{j'}w_{z1j'}\frac{(1-s_1)}{s_{j'}} + \delta_{j'}w_{z2j'}[\frac{z_1(1-s_1)}{s_{j'}} + \frac{(1-s_1)}{(\partial v_1/\partial z_1)(1-2s_1)s_{j'}}]\right)}{\frac{\partial v_1}{\partial z_1}(1-2s_1)s_1 s_{j'}\left(-\frac{\partial \tilde{v}_{j'}}{\partial x} + \gamma_{j'}w_{x1j'}\frac{(1-s_1)}{s_{j'}} + \delta_{j'}w_{x2j'}[\frac{z_1(1-s_1)}{s_{j'}} + \frac{(1-s_1)}{(\partial v_1/\partial z_1)(1-2s_1)s_{j'}}]\right)} \\
&= \frac{\frac{\partial v_1}{\partial z_1}(1-2s_1)s_1 s_{j'}\left(-\frac{\partial \tilde{v}_{j'}}{\partial z} + \gamma_{j'} + \delta_{j'}\right)}{\frac{\partial v_1}{\partial z_1}(1-2s_1)s_1 s_{j'}\left(-\frac{\partial \tilde{v}_{j'}}{\partial x} + \gamma_{2j'} + \delta_{2j'}\right)} \\
&= \frac{-\frac{\partial \tilde{v}_{j'}}{\partial z} + \gamma_{j'} + \delta_{j'}}{-\frac{\partial \tilde{v}_{j'}}{\partial x} + \gamma_{2j'} + \delta_{2j'}}
\end{aligned} \tag{50}$$

Thus, we have:

$$\frac{\partial^2 s_1}{\partial z_1 \partial z_{j'}} / \frac{\partial^2 s_1}{\partial z_1 \partial x_{j'}} = \frac{-\frac{\partial \tilde{v}_{j'}}{\partial z} + \gamma_{j'} + \delta_{j'}}{-\frac{\partial \tilde{v}_{j'}}{\partial x} + \gamma_{2j'} + \delta_{2j'}} \tag{51}$$

where $w_{z1j'} = w_{x1j'} = \frac{s_{j'}}{1-s_1}$ and $w_{x2j'} = w_{z2j'} = (\partial v_1/\partial z_1)\frac{(1-2s_1)s_{j'}}{1-s_1}(1 + (\partial v_1/\partial z_1)(1-2s_1)z_1)^{-1}$. Given a linear specification of $\tilde{v}$, $\tilde{v}(x_j, z_j) = x_j a_1 + z_j a_2$, 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 naive 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 (17) is estimated treating these weights as constants.[49]

To recover estimates of β/α from the flexible logit model, we use the ratio in equation (51). With the linear specification of $\tilde{v}$, this ratio is given by $\frac{\beta}{\alpha} = \frac{-a_2+\gamma_{j'}+\delta_{j'}}{-a_1+\gamma_{2j'}+\delta_{2j'}}$. 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 $\gamma_k = \gamma$, $\gamma_{2k} = \gamma_2$, $\delta_k = \delta$ and $\delta_{2k} = \delta_2$, which gives a single estimate of $\frac{\beta}{\alpha}$.

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

Suppose that utility is given by $U_{ij} = x_j\alpha + z_j\beta + \epsilon_{ij}$ and that consumers search sequentially according to the model of Weitzman (1979).

As shown in Armstrong (2017),[50] 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

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

[49] Since $\partial v_1/\partial z_1$ is estimated imprecisely from the naive 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.

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

we can write

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

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

$$\frac{\partial s_1}{\partial z_1} = \left[\frac{\partial}{\partial z_1} P\left(\epsilon_k - \epsilon_1 \leq \left(x_1 - x_k\right)\alpha + \left(z_1 - z_k\right)\beta \ \forall k\right)\right] P\left(rv \geq z_1\right) \tag{52}$$

Given identification of (α, β) by the argument in Section 2.1, the first term on the rhs of (52) is identified given parametric assumptions on the distribution of ϵ. 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,[51] the distribution of c is also identified.

Equation (52) also lends itself to a different argument that does not require making a parametric assumption on the distribution of ϵ, but instead relies on "at-infinity" variation. Note that the first term on the rhs of (52) is invariant to increasing all z_j's by the same amount. Thus, we can write

$$\frac{\frac{\partial s_1}{\partial z_1}\left(\mathbf{z} + \mathbf{\Delta}\right)}{\frac{\partial s_1}{\partial z_1}\left(\mathbf{z}\right)} = \frac{P\left(rv \geq z_1 + \Delta\right)}{P\left(rv \geq z_1\right)} \tag{53}$$

where $\mathbf{\Delta}$ is a $J-$vector with all elements equal to some Δ. Letting $\Delta \to -\infty$, the numerator on the rhs of (53) 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.

Appendix E: Welfare Benefits of Information

Appendix D of Abaluck and Gruber (2009) shows that dollar-equivalent consumer surplus in logit models where positive preferences (i.e., preferences describing potentially uninformed behavior) are given by β_{pos} and normative preferences (i.e., those relevant for welfare evaluations) are given by

[51] 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.

β_{norm} can be computed as:

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

where α_p is the (normative) marginal utility of income, estimated as the coefficient on price. Once consumers are informed and their preferences are β_{norm}, consumer surplus is given by the conventional log-sum formula:

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

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

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