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

Jason Abaluck and Giovanni Compiani*

November 20, 2019

Abstract

We state a sufficient condition under which choice data alone suffices to identify consumer preferences when choices are not fully informed. Suppose that: (i) the data generating process is a search model in which the attribute hidden to consumers is observed by the econometrician; (ii) if a consumer searches good j, then she also searches goods which are better than j in terms of the non-hidden component of utility; and (iii) consumers choose the good that maximizes overall utility among searched goods. Canonical models will be biased: the value of the hidden attribute will be understated because consumers will be unresponsive to variation in the attribute for goods that they do not search. Under the conditions above and additional mild restrictions, an alternative method of recovering preferences using cross derivatives of choice probabilities succeeds regardless of the search protocol and is thus robust to what consumers know when they choose. The approach nests several standard models, including full information. In addition, it requires few assumptions on utility beyond those imposed in the existing literature on discrete choice demand. Our approach can be used to recover preferences when consumers are imperfectly informed as well as to forecast how consumers will respond to information.

*Thanks to Abi Adams, Judy Chevalier, Magne Mogstad, Aniko Oery, Nicholas Ryan, Fiona Scott Morton, Raluca Ursu, Miguel Villas-Boas, and seminar participants at Yale, SICS 2019, and Marketing Science 2019 for helpful comments and suggestions. Tianyu Han and Jaewon Lee provided excellent research assistance.

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 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. 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", whether their prior beliefs are systematically biased, 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)).

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 and without specifying the full structural search model. This condition relies on what we call *visible utility*, the component of utility visible prior to search. The main restriction we impose on search is that consumers never search items with lower visible utility and then leave unsearched items with higher visible utility (a condition we make precise below). We show that if this condition is satisfied, there is a function of choice probabilities which recovers preferences whether consumers are fully or partially informed. Given preferences recovered by our approach, we show that it is possible to identify other primitives of interests. For example, in a sequential search model à la Weitzman (1979), one can identify the distribution of search costs and thus simulate counterfactuals in which search costs are altered by information provision or other means. However, preferences are identified—and thus can be estimated—without imposing a structural search model beyond the visible utility assumption. Specifically, no additional assumptions about option value, optimization vs. satisficing, rational expectations, simultaneous or sequential search, or distributional assumptions about beliefs and preferences are necessary for identification.

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 also 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 has 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 $U_{ij} = x_j\alpha + z_j\beta + \epsilon_{ij}$ where (without loss) $\alpha > 0$ and $\beta > 0$. In the text, we extend this result to allow vector-valued x_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: $\left|\frac{\partial s_j}{\partial z_j} / \frac{\partial s_j}{\partial x_j}\right| < |\beta/\alpha|$.[1] 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. Visible utility has the property that if $VU_{ij} \geq VU_{ij'}$ and $z_j \geq z_{j'}$, then $U_{ij} \geq U_{ij'}$. Given this definition, suppose consumer search is characterized by the following assumptions:

1. If i searches a good with visible utility VU_{ij}, she also searches all goods j' with $VU_{ij'} \geq VU_{ij}$;

2. Conditional on searching j, consumer i searches j' if and only if $g_i(x_{j'}, \epsilon_{ij'}, U_{ij}) \geq 0$ where g_i is a monotonically decreasing function of U_{ij};

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

The first assumption states that if you search a good with a given level of visible utility, you always search all goods with higher visible utility. This assumption would be satisfied if consumers searched in order of visible utility but it is weaker (e.g. consumers might search the second ranked good in terms of visible utility and then the first, provided they eventually search both). The second assumption states that you

[1] We prove this in Section 2.1.

are less likely to search if utility in hand U_{ij} is larger and that your decision to search the next good only depends on x_j and z_j via U_{ij}. This rules out models where, for example, you stop searching when you find a good with large enough z_j regardless of U_{ij}. The third assumption simply says that you maximize utility conditional on search.

These assumptions are 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 with the highest visible utility and decide whether to search the next good by comparing the expected benefits with search costs. These assumptions are also consistent with many behavioral models: 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 search all goods with visible utility above a certain threshold and then give up. One important special case is when consumers search all goods, yielding a conventional full-information discrete choice model.

Our main result (for the case with two goods, linear utility and no random coefficients) is that, if the above assumptions are met, then $\frac{\partial^2 s_1}{\partial z_1 \partial z_2} / \frac{\partial^2 s_1}{\partial z_1 \partial x_2} = \beta/\alpha$, 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 case. The intuition is as follows: under our assumptions, consumers who search good 1 will always choose the good that maximizes utility among all options (although they might not be fully informed).[2] Therefore, z_1 only impacts choices for consumers who maximize utility. By examining how $\frac{\partial s_1}{\partial z_1}$ changes with attributes of rival goods, we are effectively learning about the impact of z_2 and x_2 on choices for consumers who act *as if* they were fully informed. For these consumers, both of these attributes impact choices only via U_2, and they do so in proportion to β/α. Since α is straightforwardly identified by examining choices where $z_j = z$ for all j,[3] this means that preferences are identified. The quantity $\frac{\partial^2 s_1}{\partial z_1 \partial z_2} / \frac{\partial^2 s_1}{\partial z_1 \partial x_2}$ thus provides robust estimates of marginal rates of substitution in the sense that it works regardless of the search process that consumers engage in provided the assumptions above are met.

How general is this result? Using additional derivatives of the share 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 $U_{ij} = h(x_j \alpha_i + z_j \beta_i) + \epsilon_{ij}$ for a known function h, 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, the approach allows for unobserved heterogeneity as flexibly as any constructive results in the existing literature on discrete choice demand.[4]

[2]Again, if some other good had higher utility than good 1, that good must have higher visible utility since good 1 has the best z.

[3]We show this formally in Section 2.

[4]Fox and Gandhi (2016) provide identification results for more general models allowing for both nonlinearity and flexible

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. For example, when the number of goods or attributes is not too large, one could approximate the share functions via the method of sieves, which makes it possible to conveniently impose natural restrictions such as monotonicity and exchangeability.[5] When the number of goods is too large for nonparametric estimation, we suggest using "flexible logits", which allow attributes of rival goods to directly enter utility for each alternative and thus flexibly parameterize the cross-derivatives that identify preferences in our model.

Our result relates to several existing literatures. A large theoretical and empirical literature investigates the formation of "consideration sets" (e.g. Roberts and Lattin (1991), Conlon and Mortimer (2013), Goeree (2008), and Gaynor, Propper, and Seiler (2016)). 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. This paper considers the complementary problem of imperfect information at the level of attributes rather than goods. The recent theoretical literature on this question includes Branco, Sun, and Villas-Boas (2012) and Ke, Shen, and Villas-Boas (2016). A few papers, such as Mehta, Rajiv, and Srinivasan (2003) and Honka and Chintagunta (2016) consider estimating utility given explicit search models and attempt to recover preferences conditional on search. 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.[6]

An important 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.

---

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), which we utilize in our proof. The conditions required differ slightly from Fox, Kim, Ryan, and Bajari (2012): they recover distributions satisfying the "Carleman condition," which implies that the distribution of preferences is uniquely characterized by its moments. Alternatively, we recover arbitrary weights for density functions with support on a fixed grid. 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.

[5]Compiani (2019) proposes a sieve approach to estimate demand for differentiated products based on aggregate data.

[6]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.

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 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. Our paper can be thought of as a data driven way of identifying a subgroup of consumers who maximize utility (those who search the good with the highest value of the hidden attribute). Importantly, the subgroup we identify is *not necessarily informed* in the sense of searching all available goods. What we know is that they maximize utility among all options, meaning that their choices can be used to estimate preferences even though they do not themselves know that they do so, as they cannot verify that they have searched the good with the highest z_j without searching all goods. This insight—that we can recover preferences by identifying consumers who maximize utility rather than those who are informed—is what makes possible the surprising recovery of preferences from choices in cross-sectional data without auxiliary information.

Section 2 lays out our formal framework and proves our identification results, Section 4 provides details of estimation and simulation results, and Section 5 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.[7,8] For simplicity, we assume that x_j is scalar for all j, but the identification argument immediately extends to the case of vector-valued x_j's. Without loss of generality, we order the goods by z_j, with good 1 having the largest value of z_j. Again, for simplicity, we focus on the case where z_j is scalar-valued. If there are multiple hidden attributes for each good j, call them z_{kj}, the same argument applies with good 1 defined as the good satisfying $z_{k1} \geq z_{kj}$ for all j and k.[9] 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 both x_j and z_j (and without loss, we

[7]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.

[8]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.

[9]Note that, while in the scalar z_j case a good satisfying our definition of good 1 always exists (we can always rank options based on their z_j's), this is not true of the vector-valued z_j case. In this sense, the argument for the vector-valued z_j case is more demanding of the data.

assume it is increasing).

(*ii*) For all i, the function $b_{ij}(x_j, z_j)$ is not alternative-specific, i.e. $b_{ij}(x_j, z_j) = b_i(x_j, z_j)$.

(*iii*) For all i and j, a_{ij} and b_i are infinitely differentiable.

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[10] 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 $\tilde{a}_{ij}(x_j) = \alpha_i x_j$, $a_{ij}(x_j) = \alpha_i x_j + \epsilon_{ij}$, and $b_i(x_j, z_j) = \beta_i z_j$. 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 $\tilde{a}_{ij}(x_j) = \alpha x_j$, $a_{ij}(x_j) = \alpha x_j + \epsilon_{ij}$, 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*) If consumer i searches a good with visible utility VU_{ij}, she also searches all goods j' with $VU_{ij'} \geq VU_{ij}$;

(*ii*) Consumer i searches j' if and only if, for all j that i has searched so far, $g_i(x_{j'}, \epsilon_{ij'}, U_{ij}) \geq 0$ where g_i is a monotonically decreasing function of U_{ij} and is infinitely differentiable in its first and third arguments;

(*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.

We discuss these conditions at length in Section 2.4. To briefly clarify, Assumption 2(*i*) states that if you search a good j, you always search all goods with higher visible utility than j. This would be satisfied if consumers searched in order of visible utility but it is weaker (e.g. consumers might search the second ranked good in terms of visible utility and then the first, provided they eventually search both). Assumption 2(*ii*) states that consumers decide whether or not to keep searching based on the utility in hand and the visible utility of the good they are considering searching. This rules out, for example, a search protocol whereby one stops searching after discovering a large z_j irrespective of utility in hand. We subscript the stopping rule 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.

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 r'_{ij} defined implicitly by*

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

[10] We allow for nonlinearities subject to Assumptions 1(*i*) and 1(*iii*) being satisfied.

where $r_i \equiv \frac{r'_{ij} - \alpha_i x_j - \epsilon_{ij}}{\beta_i}$ *and the last steps follows from a change of variable. We can interpret* r_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 is lower than* $x_{j'} \alpha_i + r_i \beta_i + \epsilon_{ij'}$*. Once she stops searching, she maximizes utility among the searched goods. Thus, Assumption 2 is satisfied with* $g_i\left(x_{j'}, \epsilon_{ij'}, U_{ij}\right) = x_{j'} \alpha_i + r_i \beta_i + \epsilon_{ij'} - U_{ij}$*.*

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*[11] *and myopically checking whether searching the next good is worth the cost. This search protocol is accommodated by Assumption 2 with* $g_i\left(x_{j'}, \epsilon_{ij'}, U_{ij}\right) = E_z \max\{U_{ij'}, U_{ij}\} - U_{ij} - c_i$ *where* c_i *are individual-specific search costs.*

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* $g_i\left(x_{j'}, \epsilon_{ij'}, U_{ij}\right) = \tau_i - U_{ij}$*.*

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

Example 5. *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* $g_i\left(x_{j'}, \epsilon_{ij'}, U_{ij}\right) = \alpha x_{j'} + \epsilon_{ij'} - \tilde{\tau}_i$*.*

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.

We now state and prove a lemma that is at the core of our results. Let $\mathcal{G}_i$ denote the set of searched goods for individual i.

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

Proof. Recall that good 1 is defined as the good with the highest value of z_j (a quantity observed by the econometrician). If good 1 is searched but utility is not maximized, then for some unsearched j, $U_{ij} \geq U_{i1}$. Since $z_1 \geq z_j$, by monotonicity, $b_i(x, z_1) \geq b_i(x, z_j)$. By the contintuity of b_i, this implies that for η sufficiently small, $b_i(x_1, z_1) \geq b_i(x_j, z_j)$. Given this, $U_{ij} \geq U_{i1}$ implies $VU_{ij} \geq VU_{i1}$. But by Assumption 2(i), this implies that good j is searched, which is a contradiction. □

Note that this 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

[11]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.

1 or not) and thus the consumer will choose that good. Note that 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.

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 ($1 \in \mathcal{G}_i$), then i chooses the utility-maximizing good.*

Proof. If good 1 is searched but utility is not maximized, then for some unsearched j, $U_{ij} \geq U_{i1}$. Since $z_1 \geq z_k$ for all k, it must be that $x_j\alpha_i + \epsilon_{ij} \geq x_1\alpha_i + \epsilon_{i1}$. 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, we need one more definition. We define choice probabilities s_j as:

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

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 $\mathbf{z} \equiv [z_1, \cdots, z_J]$ (as well as the unknown parameters determining the search process), but we 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 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*. 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 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 $U_{ijt} = v_i(x_j, z_j) + \epsilon_{ijt}$

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

These cases subsume the most general possible models in which constructive identification results for structural parameters exist in discrete choice contexts.

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

Theorem 1. *Let Assumption 2 hold and utility be given by $U_{ij} = v(x_j, z_j) + \epsilon_{ij}$ where the function v and the cdf of $\epsilon_i \equiv (\epsilon_{i1}, \epsilon_{i2})$ are infinitely differentiable and ϵ_i is supported on $\mathbb{R}^2$. Further, assume that the function g is also infinitely differentiable. Then, $\frac{\partial^n v}{\partial z^{\bar{n}} \partial x^{n-\bar{n}}}(0,0)$ is identified for all $n \geq 1$, $\bar{n} \leq n$, and thus v is recoverable with arbitrary precision 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 we impose the requirement that the function v be infinitely differentiable. If we put more restrictions on the utility function, then we can recover preferences under much weaker differentiability requirements.

Corollary 1. *Let the assumptions of Theorem 1 be satisfied. Then, the marginal rates of substitution can be recovered using:*

$$\frac{\partial^2 s_1}{\partial z_1 \partial z_2} \Big/ \frac{\partial^2 s_1}{\partial z_1 \partial x_2} = \frac{\partial v}{\partial z_2} \Big/ \frac{\partial v}{\partial x_2} \tag{4}$$

If v is linear, this suffices to recover preferences given identification of $\frac{\partial v}{\partial x_2}$, which we show below.

While we state this result as a corollary of Theorem 1, it holds under weaker conditions. In particular, as can be seen from (4), marginal rates of substitution may be recovered by taking ratios of second derivatives of the share function for good 1. Therefore, this result only requires twice differentiability of the cdf of ϵ_i and of the functions v and g_i.

In Appendix A.1, we provide a general proof of Theorem 1 and Corollary 1 which allows for nonlinear utility and $J \geq 2$. Here, we focus on the linear case with $J = 2$ in order to highlight the core idea of Theorem 1 in a simple setting. In this case, with $U_{ij} = x_j\alpha + z_j\beta + \epsilon_{ij}$, the theorem (and Corollary 1) imply that:

$$\frac{\partial^2 s_1}{\partial z_1 \partial z_2} \Big/ \frac{\partial^2 s_1}{\partial z_1 \partial x_2} = \frac{\beta}{\alpha} \tag{5}$$

This suffices to recover preferences, since 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.

Proof. In order to ease notation, we often suppress the subscript i in what follows. In the two-good case with linear utility, visible utility is given by $x_j\alpha + \epsilon_{ij}$. As above, good 1 is defined as the good with the highest value of z_j. The probability of choosing good 1 can be written as:

$$\begin{aligned} s_1 &= P(\mathcal{G}=1) + P\left(\{U_1 > U_2\} \cap \{\mathcal{G}=\{1,2\}\}\right) \\ &= P(U_1 > U_2) - P\left(\{U_1 > U_2\} \cap \{\mathcal{G}=2\}\right) \end{aligned} \tag{6}$$

where, as above, $\mathcal{G}$ denotes the set of searched goods. The second line 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 will be searched first and chosen).[12]

Our goal will be to show that both z_2 and x_2 only impact $\frac{\partial s_1}{\partial z_1}$ via U_2. This, in turn, implies that $\frac{\partial^2 s_1}{\partial z_1 \partial z_2} = \frac{\partial^2 s_1}{\partial z_1 \partial U_2}\frac{\partial U_2}{\partial z_2}$ and $\frac{\partial^2 s_1}{\partial z_1 \partial x_2} = \frac{\partial^2 s_1}{\partial z_1 \partial U_2}\frac{\partial U_2}{\partial x_2}$, and the result in equation (5) follows. (Note that $\frac{\partial^2 s_1}{\partial z_1 \partial x_2} \neq 0$ and thus we can divide by it due to the full support assumption on ϵ_i.)

To establish this, note that we can write:

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

where the second line follows since $VU_1 > VU_2$ implies $U_1 > U_2$ and thus $P\left(\{VU_1 > VU_2\} \cap \{g(x_1, \epsilon_1, U_2) \leq 0\}\right) = P\left(\{U_1 > U_2\} \cap \{VU_1 > VU_2\} \cap \{g(x_1, \epsilon_1, U_2) \leq 0\}\right)$. The second term in equation (7) is not a function of z_1. The first term is only a function of x_2 and z_2 via 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 U_2, thus proving the theorem. □

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 simplicity, consider the model with linear utility $U_{ij} = x_j\alpha + z_j\beta + \epsilon_{ij}$ and assume that the function g_i in Assumption 2(*ii*) is increasing in $x_{j'}\alpha$. This condition is satisfied in all the search model considered above (Examples 1–4) 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. Define $v_j \equiv x_j\alpha + z_j\beta$, $\mathbf{v} = (v_1, v_2)$ and let

$$P_2^*(\mathbf{v}, \mathbf{x}) \equiv P(\{U_1 > U_2\} \cap \{VU_2 > VU_1\} \cap \{g(x_1, \epsilon_1, U_2) \leq 0\})$$

Plugging equation (7) into equation (6) and differentiating, we obtain:

$$\begin{aligned} \frac{\partial s_1}{\partial z_1} &= \beta \left[\frac{\partial P(U_1 > U_2)}{\partial v_1} - \frac{\partial P_2^*}{\partial v_1}(\mathbf{v}, \mathbf{x})\right] \\ \frac{\partial s_1}{\partial x_1} &= \alpha \left[\frac{\partial P(U_1 > U_2)}{\partial v_1} - \frac{\partial P_2^*}{\partial v_1}(\mathbf{v}, \mathbf{x}) - \frac{1}{\alpha}\frac{\partial P_2^*}{\partial x_1}(\mathbf{v}, \mathbf{x})\right] \end{aligned}$$

[12]Formally, $P(U_1 > U_2) = P\left(\{U_1 > U_2\} \cap \{\mathcal{G}=1\}\right) + P\left(\{U_1 > U_2\} \cap \{\mathcal{G}=2\}\right) + P\left(\{U_1 > U_2\} \cap \{\mathcal{G}=\{1,2\}\}\right)$. Since searching good 1 first implies that good 1 has higher $x_j\alpha + \epsilon_{ij}$ and good 1 has higher z_j by definition, we have $P\left(\{U_1 > U_2\} \cap \{\mathcal{G}=1\}\right) = P(\mathcal{G}=1)$, and the result in equation (6) follows.

Note that $\frac{\partial P(U_1>U_2)}{\partial v_1}-\frac{\partial P_2^*}{\partial v_1}(\mathbf{v},\mathbf{x})=\frac{\partial P(\{U_1>U_2\}\cap\{g(x_1,\epsilon_1,U_2)\geq 0\})}{\partial v_1}>0$. Further, $\frac{1}{\alpha}\frac{\partial P_2^*}{\partial x_1}(\mathbf{v},\mathbf{x})\leq 0$ if $\frac{\partial g}{\partial x_1}(x_1,\epsilon_1,U_2)\geq 0$.[13] Therefore,

$$\frac{\frac{\partial s_1}{\partial z_1}}{\frac{\partial s_1}{\partial x_1}}\leq\frac{\beta}{\alpha}$$

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.[14] 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 time. Given these, the following result holds:

Theorem 2. *Let Assumption 2 hold and utility be given by $U_{ijt}=v_i(x_{jt},z_{jt})+\epsilon_{ijt}$ where the functions v_i and the cdf of $\epsilon_{it}\equiv(\epsilon_{i1t},\epsilon_{i2t})$ are infinitely differentiable and ϵ_{it} is supported on $\mathbb{R}^2$. Further, assume that the function g is also infinitely differentiable. Then, $\frac{\partial^n v_i}{\partial z^{\bar{n}}\partial x^{n-\bar{n}}}(0,0)$ are identified for all $n\geq 1$, $\bar{n}\leq n$, and thus v_i is recoverable with arbitrary precision up to an additive constant.*

Corollary 2. *Let the assumptions of Theorem 2 hold. Then, marginal rates of substitution can be recovered using:*

$$\frac{\partial^2 s_{i1}}{\partial z_1\partial z_2}\Big/\frac{\partial^2 s_{i1}}{\partial z_1\partial x_2}=\frac{\partial v_i}{\partial z_2}\Big/\frac{\partial v_i}{\partial x_2}\tag{8}$$

If v is linear, this suffices to recover preferences.

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. What if we have non-separable heterogeneity? The canonical case 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 positive 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 $\tilde{\pi}_{k_\alpha,k_\beta}\equiv P\left(\{\alpha_i=\alpha_{k_\alpha}\}\cap\left\{\beta_i=\beta_{k_\beta}\right\}\right)$.

(iii) The distribution of ϵ_i is known (or independently identified) and has a $K_\alpha K_\beta-$time differentiable cdf.

Assumption 3(ii) follows Fox, Kim, and Yang (2016), and our identification proof uses techniques developed in Fox, Kim, Ryan, and Bajari (2012). Assumption 3(iii) is maintained in both of these papers.

[13] Recall that we assume $\alpha>0$ without loss.

[14] Again, here we focus on the $J=2$ case for simplicity. See Appendix A.2 for a proof for the more general case with $J\geq 2$.

Theorem 3. *Let Assumptions 2 and 3 hold. 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.2. □

2.4 Discussion of Search Model Assumptions

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

1. If consumer i searches a good with visible utility VU_{ij}, she also searches all goods j' with $VU_{ij'} \geq VU_{ij}$;

2. Consumer i searches j' if and only if, for all goods j that i has searched so far, $g_i(x_{j'}, \epsilon_{ij'}, U_{ij}) \geq 0$, where g_i is a monotonically decreasing function of U_{ij} and is infinitely differentiable in its first and third arguments;

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

4. Only the value of z_j is unknown prior to search.

As discussed above, a sufficient condition for the first assumption is that consumers search in order of visible utility. In the Weitzman (1979) search model, consumers search goods in order of reservation utility, which is a function both of the visible attributes of those goods and the distribution of the hidden attribute z_j. If z_j is i.i.d. across goods and each individual has the same search cost across all goods, then it follows that consumers will search in order of visible utility.[15] There are 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. For instance, if consumers shopping for laptops need to pay a cost to learn about hard disk space and know that hard disk space varies more across Apple laptops than Dell laptops, then they might prefer searching an Apple laptop before a Dell laptop with a higher observable utility since searching the former reveals more valuable information. Second, unobservables might be correlated across goods, and 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. Since searching in order of visible utility is a sufficient, but not necessary condition, we allow for each of these three possibilities subject to Assumption 2(i) being satisfied.

Further, note that while i.i.d. priors and search costs is one rationale for the assumption that consumers search in order of visible utility, there are many others. 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). Additionally, even if consumers are fully rational, the strength of the visible utility assumption is mitigated by the fact that we can often include in the model

[15]See Example 1.

observables which weaken the assumption of i.i.d. priors for z_j. Searching an Apple laptop may reveal information about the hard disk space for other Apple laptops, but if the observables in our model include brand fixed effects, then this is part of visible utility and not a reason for consumers not to search in order of 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 Ursu (2018). In Section 3.1, we extend our main result to these cases where some observable attributes impact search probability but not utility (thus violating the visible utility assumption stated above).

Next, the assumption that consumers search good j' if and only if $g_i(x_{j'}, \epsilon_{ij'}, U_{ij}) \geq 0$ for all goods j searched so far is natural and satisfied in many search models used in the literature, including Weitzman search, satisficing, searching all goods with visible utility above a threshold, random search, and directed cognition.

Assumption $2(ii)$ 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. To the degree that search can plausibly be modeled as binary, the first assumption is relatively innocuous. In cases where the hidden attribute is not easily available upon search, binary search may be implausible. For example, if the hidden attribute is "school-value added," a consumer who searches more may learn about test scores and graduation rates, but these are signals of the underlying variable. Our model is more suitable for cases where the covariates included in the model are ones that consumers can observe directly given search. 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.

Perhaps the most stringent assumption in our model is Assumption $2(iv)$ 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"). Our model does not accommodate cases where search reveals information not available to the econometrician. To the degree that this is a concern, one should try to code all information that may be revealed to the consumer upon search as attributes in the model. The viability of this strategy will of course vary from case to case.

2.5 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 information of this type: 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 resorting to the assumptions we require here.

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.[16]

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. In the linear case, visible utility is given by:

$$VU_{ij} = x_j\alpha_i + \epsilon_{ij} \tag{9}$$

In our proof, examining choices with equal values of the hidden attribute is sufficient to identify the distribution of α_i. This result nonetheless relies on the visible utility assumption (in general cases, even if z_j is equal for all goods, consumers do not necessarily maximize $x_j\alpha_i + \epsilon_{ij}$ because they may not search all goods). Alternative assumptions may suffice 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:

$$P(j \in \mathcal{G}_i|\mathbf{x}) = \sum_k P(|\mathcal{G}_i| = k|\mathbf{x})P(j \in \mathcal{G}_i||\mathcal{G}_i| = k, \mathbf{x}) \tag{10}$$

since the first probability on the rhs is observed and the second is pinned down by the model assumptions. Checking (10) 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

[16] 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.

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:

$$s_j(x) \leq 1 - P(U_{ik} > U_{ij} \text{ and } VU_{ik} > VU_{ij} \text{ for some k}) \tag{11}$$

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:

$$s_j(x) \geq P(U_{ij} > U_{ik} \text{ and } VU_{ij} > VU_{ik} \text{ for all k}) \tag{12}$$

Once again, this probability can be computed given knowledge of preferences and the distribution of ϵ.

2.6 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, this type of large support assumptions are restrictive and in practice leverage only a small fraction of the data, which is problematic in estimation.

In contrast, our proof requires much more plausible support assumptions. There is always a good which maximizes z_j. 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, requiring that we can estimate many higher order derivatives of choice probabilities. We discuss these challenges further in Section 4.

3 Extensions

In this section, we consider three extensions to the baseline model: the first two relax the visible utility assumption, while the third deals with the case of endogenous attributes.

3.1 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.

Our model 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 in order of VU_{ij}, we assume that goods are searched in order of $m(VU_{ij}, r_j)$ where $m(\cdot)$ is strictly increasing in both VU_{ij} and r_j.

We formalize this, together with two additional restrictions, in the following assumption.

Assumption 4. (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 $m(\cdot)$ is strictly increasing in both arguments;

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

(iii) r_j is unrelated to unobservable components of utility, and locally perturbing x and z does not perturb r_j.

Assumption 4(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:

Lemma 3. *Let Assumptions 1, 2(iii)-2(iv), and 4 hold and let $x_j \in [x - \eta, x + \eta]$ for all j, for some $\eta > 0$ sufficiently small. If consumer i searches good 1 (i.e. $1 \in \mathcal{G}_i$), then i chooses the good which maximizes utility among all goods with $r_j \geq r_1$.*

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.

Let R denote the set of goods j with $r_j \geq r_1$ (R includes at least two goods by assumption 4(ii)). Then, our previous arguments still go through provided we define everything from the set of goods in R rather than all goods and R_1 as $R \setminus 1$. To fix ideas, consider the case with two goods (where to satisfy the above assumptions, we must have $r_2 > r_1$). Then we have:

$$\begin{aligned} s_1 &= P(U_1 > U_2) - P(U_1 > U_2, m(VU_2, r_2) > m(VU_1, r_1), g_i(x_1, \epsilon_1, U_2) < 0) \\ &= P(U_1 > U_2) - P(U_1 > U_2, VU_2 > VU_1, m(VU_2, r_2) > m(VU_1, r_1), g_i(x_1, \epsilon_1, U_2) < 0) \\ &\quad -P(U_1 > U_2, VU_1 > VU_2, m(VU_2, r_2) > m(VU_1, r_1), g_i(x_1, \epsilon_1, U_2) < 0) \\ &= P(U_1 > U_2) - P(U_1 > U_2, VU_2 > VU_1, g_i(x_1, \epsilon_1, U_2) < 0) \\ &\quad -P(VU_1 > VU_2, m(VU_2, r_2) > m(VU_1, r_1), g_i(x_1, \epsilon_1, U_2) < 0) \end{aligned} \tag{13}$$

The third equality uses the fact that $VU_2 > VU_1$ and $r_2 > r_1$ implies $m(VU_2, r_2) > m(VU_1, r_1)$ by monotonicity, and $VU_1 > VU_2$ implies $U_1 > U_2$ since $z_1 > z_2$. The second term in equation (13) is the object from our usual proof that can be separated into a component which only depends on x_2 via U_2 and a component which is not a function of z_1. The third term is likewise not a function of z_1. Thus, our argument still holds.

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

Another reason why the visible utility assumption $2(i)$ might fail is that consumers might form expectations on 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 order to accommodate this type of pattern, we consider the linear model $U_{ij} = x_j\alpha + z_j\beta + \epsilon_{ij}$ and re-write it as

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

where we assume that consumers use the linear projection $\mathbb{E}(z_j|x_j) = \gamma_0 + \gamma_1 x_j$ and let $\tilde{z}_j \equiv z_j - \mathbb{E}(z_j|x_j)$. Visible utility is then given by $\beta\gamma_0 + x_j(\alpha + \beta\gamma_1)$ and consumers learn the deviation from their expectation on z_j, $\tilde{z}_j$, upon searching. We also assume that consumers' expectations are correct, i.e. that the coefficients γ_0 and γ_1 can be recovered by regressing z_j on x_j.

Then, Corollary (1) yields identification of $\frac{\beta}{\alpha+\beta\gamma_1}$. Finally, we show that $\alpha + \beta\gamma_1$ can be recovered, so that α and β are separately identified. 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

$z_j + \epsilon(\delta) - \mathbb{E}(z_j | x_j + \delta) = 0$. 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

$$\frac{s_j(x_j + \delta, \mathbf{x}_{-j}, z_j + \epsilon(\delta), \mathbf{z}_{-j}) - s(\mathbf{x}, \mathbf{z})}{\delta} \tag{14}$$

$$= \frac{P((x_j + \delta)(\alpha + \beta\gamma_1) + \epsilon_{ij} \geq x_k(\alpha + \beta\gamma_1) + \epsilon_{ik} \forall k) - P(x_j(\alpha + \beta\gamma_1) + \epsilon_{ij} \geq x_k(\alpha + \beta\gamma_1) + \epsilon_{ik} \forall k)}{\delta}$$

$$\xrightarrow{\delta \to 0} \frac{\partial}{\partial VU_j} P(VU_{ij} \geq VU_{ik} \forall k)(\alpha + \beta\gamma_1) \tag{15}$$

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 (14) is known for all $\delta > 0$; and (ii) at $\mathbf{x} = \mathbf{0}$, the term $\frac{\partial}{\partial VU_j} P(VU_{ij} \geq VU_{ik} \forall k)$ is known if the distribution of ϵ_i is assumed to be known (e.g. extreme value). Thus, evaluating the last display at $\mathbf{x} = \mathbf{0}$ yields identification of $(\alpha + \beta\gamma_1)$.

3.3 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 consumer preferences relative to what 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 the unobservables at the product level.

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

$$u_{ij} = \alpha x_j + \beta_i z_j + \gamma_i p_j + \xi_j + \epsilon_{ij} \tag{16}$$

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.[17] If firms also know ξ_j when choosing p_j, then the two will typically be correlated, thus leading to endogeneity of p_j. In addition, we assume that p_j is also part of visible utility, so that consumers know $\alpha x_j + \gamma p_j + \xi_j + \epsilon_{ij}$ for all j prior to search and choose whether to learn z_j for any given good. This modeling choice is motivated by the fact that endogeneity of price is typically a first-order concern and prices are often available to consumers before engaging in search (e.g. one can see prices on the search results page on Amazon before clicking on any of the options).

We also modify assumption $2(ii)$ to account for the fact that p_j and ξ_j are part of the visible utility for all goods j. To this end, we let $\delta_j = \alpha x_j + \xi_j$ for all j.

[17] Note that the utility specification in (16) 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

$$u_{ij} = \tilde{\alpha}_i(\alpha x_j + \xi_j) + \beta_i z_j + \gamma_i p_j + \epsilon_{ij} \tag{17}$$

The latter is weaker, but also less common in the discrete choice literature, so we focus on model (16) in what follows.

Assumption 5. Consumer i searches j' if and only if, for all j that i has searched so far, $g_i(\delta_{j'}, p_{j'}, \epsilon_{ij'}, U_{ij}) \geq 0$ where g_i is a monotonically decreasing function of U_{ij} and is infinitely differentiable in its first and last arguments.

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'.

Letting $\delta = [\delta_1, \cdots, \delta_J]$, we may write the share of good j as

$$s_j = \sigma_j (\delta, \mathbf{z}, \mathbf{p}) \tag{18}$$

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

$$\mathbf{s} = \sigma (\delta, \mathbf{z}, \mathbf{p}) \tag{19}$$

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).[18] 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 γ_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 (18) 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

$$\mathbb{E} (\xi_j | \mathbf{x}, \mathbf{z}, \mathbf{w}) = 0 \tag{20}$$

almost surely (Assumption 3 in BH) and assume that the instruments shift the endogenous variables (market shares and endogenous attributes p_j) 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. *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.*

Proof. 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

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

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

□

4 Estimation

Our identification result shows that preferences can be recovered given knowledge of the choice probability function for good 1, denoted by $s_1(\mathbf{x}, \mathbf{z})$. In this section, we consider two approaches to estimating $s_1(\mathbf{x}, \mathbf{z})$: a nonparametric approach which is viable when the number of goods and attributes is small, and a parametric approach which can be used when the number of goods and attributes is large.

4.1 Nonparametric Estimation

When the number of goods and attributes is not too large, nonparametric methods can be used to estimate $s_1(\mathbf{x}, \mathbf{z})$. For example, one could adapt the approach in Compiani (2019) to approximate the demand function via Bernstein polynomials. This allows the researcher to impose natural restrictions on the function $s_1(\mathbf{x}, \mathbf{z})$, such as monotonicity in own- and rival attributes or exchangeability in rival attributes. In practice, such constraints may be easily enforced via linear constraints on the coefficients to be estimated. 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 inherent in nonparametric estimation.

Simulation Results Here, we present simulation results for nonparametric estimators based on Bernstein polynomial approximations to the choice probability functions. As discussed above, we impose the monotonicity restrictions that s_j be increasing in x_j and z_j and decreasing in x_k, z_k for $k \neq j$, as well as exchangeability across goods (see Compiani (2019) for a discussion of how to implement these constraints).

First, we consider a Weitzman search model. For all j, x_j and z_j are drawn from independent standard normal distributions and the shocks ϵ_{ij} are iid type I extreme value. We consider two settings: one with low search costs (drawn from a lognormal(-3,1)) and another with higher search costs (drawn from a lognormal(-1,2.5)). The coefficient α on x is set to 1 throughout. Table 1 reports the results. The degree for the polynomial approximation was chosen to be fairly low and not optimized in order to achieve

Table 1: Weitzman Search: Performance of the Bernstein polynomial nonparametric estimator

β	J	Search Costs	Bias	S.E.
1	2	Low	-0.03	0.48
1	3	Low	0.25	0.25
4	2	Low	0.19	0.12
4	3	Low	0.11	0.60
1	2	High	-0.19	0.75
1	3	High	-0.11	0.07
4	2	High	-0.25	0.62
4	3	High	-0.34	0.43

Note: Across all rows, the data is generated according to a Weitzman search model with $\alpha = 1$. The sample size is 5,000.

Table 2: Random Number of Searches: Performance of the Bernstein polynomial nonparametric estimator

β	Bias	S.E.
1	-0.18	0.41
2	-0.08	0.36
3	-0.06	0.49
4	-0.13	0.48

Note: Across all rows, the data is generated according to a model in which $\alpha = 1$ and consumers search one or two goods with equal probabilities. The sample size is 5,000.

a balance between bias and variance. As a result, the nonparametric estimator is quite biased in some designs. However, the bias is much lower than that incurred by a naive logit estimator (see Table 3 below for comparison).

Second, we consider a data generating process (dgp) with $J = 2$ goods in which half of the consumers only search the good with the highest visible utility and the remaining half search both goods. Note that this dgp in general does not satisfy our Assumption 2(ii). This is because the stopping rule is a function of how many goods one has searched so far and thus does not just depend on the visible utility of the next good under consideration, but on the entire $J-$vector of visible utilities. Therefore, considering the performance of our estimator applied to this dgp sheds some light on its robustness to this type of misspecification. Table 2 shows that the estimator is quite robust.

4.2 Parametric Estimation

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 relatively large number of goods. We focus specifically 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 $\frac{\partial^2 s_1}{\partial z_1 \partial z_j} / \frac{\partial^2 s_1}{\partial z_1 \partial x_j}$. In more general non-linear or random coefficients models, the identification arguments require recovery of higher-order derivatives and thus might not directly translate into viable estimation strategies in small to medium sample sizes. 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 previously established in the literature (see Section 4.3 and Appendix A.3 for more on this).

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 $u_{ij} = v_j^* + \epsilon_{ij}$ where $v_j^* = x_j a_1 + z_j a_2 + x_j^2 a_3 + z_j^2 a_4 + x_j z_j a_5$ and ϵ_{ij} is i.i.d. extreme value, then for $q_j \in \{x_j, z_j\}$:

$$\begin{aligned} \frac{\partial s_j}{\partial q_j} &= \frac{\partial s_j}{\partial v_j^*}\frac{\partial v_j^*}{\partial q_j} = \frac{\partial v_j^*}{\partial q_j} s_j(1-s_j) \\ \frac{\partial s_j}{\partial q_{j'}} &= \frac{\partial s_j}{\partial v_{j'}^*}\frac{\partial v_{j'}^*}{\partial q_{j'}} = -\frac{\partial v_{j'}^*}{\partial q_j'} s_j s_{j'} \\ \frac{\partial^2 s_j}{\partial z_j \partial q_{j'}} &= -\frac{\partial v_j^*}{\partial q_{j'}}\frac{\partial v_j^*}{\partial z_j} s_j s_{j'}(1-2s_j) \end{aligned} \tag{21}$$

for $j' \neq j$. Thus, in a conventional logit model, $\frac{\partial^2 s_1}{\partial z_1 \partial z_{j'}} / \frac{\partial^2 s_1}{\partial z_1 \partial x_{j'}} = \frac{\partial s_1}{\partial z_{j'}} / \frac{\partial s_1}{\partial x_{j'}} = \frac{\partial v_{j'}^*}{\partial z_{j'}} / \frac{\partial v_{j'}^*}{\partial x_{j'}}$, and when v_j is linear, this further equals $\frac{\partial s_1}{\partial z_1} / \frac{\partial s_1}{\partial x_1}$. If s_1 were generated by a search model in which consumers failed to search all goods, all three of these ratios would differ. 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 in structural search models. To allow for this additional flexibility, we let utility for good 1 depend directly on attributes of rival goods as follows:

$$v_1 = v_1^* + b_1 z_1 + \sum_{k \neq 1} \left(\gamma_k w_{z1k} z_k + \gamma_{2k} w_{x1k} x_k + w_{z2k}\delta_k z_k z_1 + w_{x2k}\delta_{2k} x_k z_1\right) \tag{22}$$

and $v_k = v_k^*$ for $k \neq 1$ where as above, $v_j^* = x_j a_1 + z_j a_2 + x_j^2 a_3 + z_j^2 a_4 + x_j z_j a_5$. 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. More precisely, the specification in equation (22) implies (treating the weights as constants):

$$\frac{\partial^2 s_1}{\partial z_1 \partial z_{j'}} / \frac{\partial^2 s_1}{\partial z_1 \partial x_{j'}} = \frac{-\frac{\partial v_{ij'}^*}{\partial z_{j'}} + \gamma_{j'} + \delta_{j'}}{-\frac{\partial v_{ij'}^*}{\partial x_{j'}} + \gamma_{2j'} + \delta_{2j'}} \tag{23}$$

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

Estimation of this model is infeasible since the levels of the choice probabilities s_1 and s_k, as well as the derivatives XXX 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 $\hat{w}_{2k}$, and then the model in equation (22) is estimated treating these weights as constants.

In our simulations, there is individual level variation in z_{ij} and x_{ij} so we could in principle compute the ratio $\frac{\partial^2 s_1}{\partial z_1 \partial z_{j'}} / \frac{\partial^2 s_1}{\partial z_1 \partial x_{j'}}$ separately for each individual as well as each alternative good j' and then average these to obtain an estimate of β/α. In practice, the parameters of the flexible logit model can only be estimated imprecisely. In our simulations, this leads to substantial bias if one attempts to average ratios in this way. To mitigate this bias, we instead suggest regressing estimates of $\frac{\partial^2 s_1}{\partial z_1 \partial z_{j'}}$ on $\frac{\partial^2 s_1}{\partial z_1 \partial x_{j'}}$. The measurement error in the estimate of $\frac{\partial^2 s_1}{\partial z_1 \partial x_{j'}}$ will still lead to some attenuation. One could mitigate this by splitting the sample, estimating the model separately in each half of the sample, and instrumenting for $\frac{\partial^2 \hat{s}_1}{\partial z_1 \partial x_{j'}}$ estimated in the first half of the sample with the same quantity estimated in the second half. In our simulations, these IV estimates give similar coefficients to the OLS regression estimates and so we report the latter.

Simulation Results To test the performance of our model, we simulate data where the underlying model is a Weitzman search model. That is, individuals are assigned search costs from a distribution we specify, they search in order of visible utility, and they sequentially decide whether to search the next good by comparing utility in hand to a reservation value which depends on search costs. In all simulations, we set $z \sim N(0, 1)$, $x \sim N(0, 4)$ and ϵ i.i.d. extreme value.

We report results from flexible logit estimation of β/α for a range of different values of β, the number of goods, and the distribution of search costs. In all cases, we set $\alpha = 1$. We consider $\beta \in \{1, 2, 3\}$, $J \in \{2, 5, 10\}$, and search costs distributed $\exp(N(-0.5, 0.5))$, $\exp(N(-1.5, 1.5))$, and $\exp(N(-1, 5))$ (respectively labeled as distributions 0, 1, and 2).

Results from these simulations are reported in Table 3. The table shows estimates of β if we estimate a conventional logit model with no adjustment for imperfect information. The coefficient is attenuated, typically biased towards zero by around 50%; there is more bias when the mean of the search cost distribution is larger. The flexible logit model is estimated by estimating equation (22) by maximum likelihood, and then regressing estimates of $\frac{\partial^2 s_1}{\partial z_1 \partial z_{j'}}$ for each individual and pair of goods $(1, j')$ on the corresponding estimates of $\frac{\partial^2 s_1}{\partial z_1 \partial x_{j'}}$. The flexible logit model does not quite recover the true values—it is just an approximation to the unknown nonparametric $s_1(x, z)$—but it performs far better than the naive logit. The bias is reduced by a factor of 5–10 across various specifications.

4.3 When Should You Estimate a Search Model?

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.

Table 3: Estimator based on cross-derivatives ratio (flexible logit) vs naive logit

β	J	Dist.	Naive Logit Est.	Cross-Derivatives Est.	Naive Logit Bias	Cross-Derivatives Bias
1	2	0	0.35	0.87	65.2%	13.5%
1	2	1	0.55	0.95	44.6%	4.8%
1	2	2	0.45	0.91	54.6%	8.6%
1	5	0	0.32	0.70	67.9%	30.2%
1	5	1	0.54	0.92	45.6%	7.6%
1	5	2	0.45	0.98	55.1%	2.3%
1	10	0	0.32	0.64	68.5%	35.5%
1	10	1	0.53	0.93	46.9%	6.9%
1	10	2	0.45	0.84	55.5%	16.4%
2	2	0	1.01	1.91	49.6%	4.5%
2	2	1	1.23	1.93	38.5%	3.4%
2	2	2	0.85	1.84	57.7%	8.1%
2	5	0	0.92	1.78	53.9%	10.9%
2	5	1	1.18	1.94	41.2%	3.2%
2	5	2	0.85	1.87	57.6%	6.4%
2	10	0	0.87	1.53	56.4%	23.6%
2	10	1	1.15	1.80	42.3%	9.8%
2	10	2	0.86	1.76	57.1%	12.0%
3	2	0	1.64	2.85	45.3%	5.0%
3	2	1	1.83	2.89	38.9%	3.5%
3	2	2	1.09	2.66	63.8%	11.2%
3	5	0	1.46	2.82	51.2%	5.9%
3	5	1	1.73	2.90	42.3%	3.5%
3	5	2	1.12	2.67	62.7%	10.9%
3	10	0	1.36	2.43	54.5%	18.9%
3	10	1	1.67	2.69	44.2%	10.5%
3	10	2	1.13	2.50	62.4%	16.6%

Note: Across all rows, the data is generated according to a Weitzman search model with $\alpha = 1$. The third column refers to different distributions for the search costs: $\exp(N(-0.5, 0.5))$, $\exp(N(-1.5, 1.5))$, and $\exp(N(-1, 5))$, labeled as 0, 1, and 2, respectively. The sample size is 5,000.

One important class of counterfactuals asks: how would consumers choose if search costs were reduced? For intermediate values of search costs, this question would require an explicit search model. On the other hand, in many cases, we might want to know how consumers would choose with *zero* search costs. This question can be answered given knowledge of preferences and available choice sets without a search model. Given the implied choice probabilities, we can also evaluate the benefits of perfect information via better choices by comparing full information choices with status quo choices.

A full normative evaluation of an information intervention would also 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. Structural search models can aid in identification of these search costs. Of course, this exercise is intrinsically challenging because such structural estimates are often sensitive to the specific assumptions made about the search process. Back of the envelope estimates of search costs based on survey data or other information on the time consumers spend choosing may be more credible than structural estimates.

Nonetheless, if search costs are of interest either to conduct counterfactuals with intermediate levels of search costs or for normative purposes, our model can be used to identify search costs given preferences and an underlying structural search model. In Appendix A.3, 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.

A final reason to estimate a full structural search model is to impose plausible 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. Our simulations suggest that in realistic samples, our results can be used to directly recover preferences in models with linear utility and homogeneous preferences over observables. However, estimation of the higher-order derivatives of choice probabilities necessary to nonparametrically identify the distribution of random coefficients 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 2.5 can be used to conduct specification tests for the internal consistency of our estimates. If these tests reject, an alternative search model may fit the data better.

5 Conclusion

We prove that it is possible to estimate preferences using only data on attributes and choices in cross-sectional data even when consumers must search to acquire information about product attributes. This result holds in a broad class of search models: those in which consumers search in order of "visible utility" and in which the searched attribute is visible to the econometrician even if unknown to consumers.

Our result allows a wide range of inquiries that are impossible using conventional methods. Prior to conducting an information intervention, choice data can be used to recover preferences and thus estimate counterfactually how consumers would choose were they fully informed. Before conducting welfare analyses, specification tests can be conducted to ask whether choices were informed. If preferences are not informed, one can recover the preferences consumers would have if they were informed.

These kinds of questions can (sometimes) be asked 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? What is the distribution of their prior beliefs? 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.

Our assumptions are sufficient for identification but not necessary. This raises several questions for future research: can our methods be extended to recover preferences when consumers do not search in order of visible utility? Likewise, can preferences be recovered if there is an "ϵ in the box:" in other words, if search reveals both attributes observable to the econometrician and observable only to consumers?

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 with welfare analyses given the data available.

Appendix A: Additional Proofs

In this appendix, we collect the proofs not included in the main text.

A.1 Proof of Theorem 1

First, we note that Assumption 2(ii), the infinite differentiability of the cdf of ϵ_i and of the function v imply that the share functions are infinitely differentiable. Next, 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 support of (X, Z) and consider a Taylor expansion of v around the point (0,0):

$$v(x,z) = v_0 + \frac{\partial v_0}{\partial x}x + \frac{\partial v_0}{\partial z}z + ... + \frac{1}{n!}\frac{\partial^n v_0}{\partial z^{\bar{n}}\partial x^{n-\bar{n}}}z^{\bar{n}}x^{n-\bar{n}} + ... \tag{24}$$

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:

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

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 $x_j \in [x - \eta, x + \eta]$ for all j 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 $\mathcal{J}_1 \equiv \{2, \cdots, J\}$, $v_j \equiv v\,(x_j, z_j)$ for all j, and $\mathbf{v} \equiv (v_1, \cdots, v_J)$. Then, by (3) we can write for all $(\mathbf{x}, \mathbf{z})$ with $z_1 \geq z_j$ for all j:

$$\begin{aligned} s_1 = P(U_1 \geq U_k\ \forall k) - \sum_{\mathcal{S}\subset\mathcal{J}_1, \mathcal{S}\neq\emptyset} & P\left(\{U_1 \geq U_k \forall k\} \cap \{VU_j \geq VU_1 \text{ for at least one } j \in \mathcal{S}\}\right. \\ & \left.\cap \{g_i\,(x_1, \epsilon_1, U_j) \leq 0 \text{ for all } j \in \mathcal{S}\} \cap \{g_i\,(x_1, \epsilon_1, U_j) \geq 0 \text{ for all } j \in \mathcal{J}_1 \backslash \mathcal{S}\}\right) \\ \equiv\ & P_4\,(\mathbf{v}) - \sum_{\mathcal{S}\subset\mathcal{J}_1, \mathcal{S}\neq\emptyset} P_5^{\mathcal{S}}\,(\mathbf{v}, x_1) \end{aligned} \tag{25}$$

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

$$\begin{aligned}
P_5^{\mathcal{S}} = \;& P\left(\{U_1 \geq U_k \; \forall k\} \cap \{g_i(x_1, \epsilon_1, U_j) \leq 0 \text{ for all } j \in \mathcal{S}\} \cap \{g_i(x_1, \epsilon_1, U_j) \geq 0 \text{ for all } j \in \mathcal{J}_1 \backslash \mathcal{S}\}\right) - \\
& P\left(\{U_1 \geq U_k \; \forall k\} \cap \{VU_1 \geq VU_j \text{ for all } j \in \mathcal{S}\}\right. \\
& \left. \cap \{g_i(x_1, \epsilon_1, U_j) \leq 0 \text{ for all } j \in \mathcal{S}\} \cap \{g_i(x_1, \epsilon_1, U_j) \geq 0 \text{ for all } j \in \mathcal{J}_1 \backslash \mathcal{S}\}\right) \\
= \;& P\left(\{U_1 \geq U_k \; \forall k\} \cap \{g_i(x_1, \epsilon_1, U_j) \leq 0 \text{ for all } j \in \mathcal{S}\} \cap \{g_i(x_1, \epsilon_1, U_j) \geq 0 \text{ for all } j \in \mathcal{J}_1 \backslash \mathcal{S}\}\right) - \\
& P\left(\{VU_1 \geq VU_j \text{ for all } j \in \mathcal{S}\} \cap \{g_i(x_1, \epsilon_1, U_j) \leq 0 \text{ for all } j \in \mathcal{S}\} \cap \{g_i(x_1, \epsilon_1, U_j) \geq 0 \text{ for all } j \in \mathcal{J}_1 \backslash \mathcal{S}\}\right) \\
\equiv \;& P_{5,1}^{\mathcal{S}}(\mathbf{v}, x_1) - P_{5,2}^{\mathcal{S}}(\mathbf{v}_{-1}, x_1)
\end{aligned}$$

where $\mathbf{v}_{-1} \equiv \mathbf{v} \setminus v_1$. 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) $g_i(x_1, \epsilon_1, U_k) \geq 0 \geq g_i(x_1, \epsilon_1, U_j)$ implies $U_k \leq U_j$, which (together with the implication in (i)) implies $U_1 \geq U_k$. Thus the event $U_1 \geq U_k \; \forall k \in \mathcal{J}_1$ is implied by the other events inside the probability and can be dropped.

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

$$\frac{\partial s_1}{\partial z_1} = \frac{\partial P_4}{\partial v_1} \frac{\partial v_1}{\partial z_1} - \sum_{\mathcal{S} \subset \mathcal{J}_1, \mathcal{S} \neq \emptyset} \frac{\partial P_{5,1}^{\mathcal{S}}}{\partial v_1} \frac{\partial v_1}{\partial z_1} \tag{26}$$

Differentiating again with respect to z_2 gives:

$$\frac{\partial^2 s_1}{\partial z_1 \partial z_2} = \frac{\partial^2 P_4}{\partial v_1 \partial v_2} \frac{\partial v_1}{\partial z_1} \frac{\partial v_2}{\partial z_2} - \sum_{\mathcal{S} \subset \mathcal{J}_1, \mathcal{S} \neq \emptyset} \frac{\partial^2 P_{5,1}^{\mathcal{S}}}{\partial v_1 \partial v_2} \frac{\partial v_1}{\partial z_1} \frac{\partial v_2}{\partial z_2} \tag{27}$$

Differentiating equation (26) with respect to x_2 gives:

$$\frac{\partial^2 s_1}{\partial z_1 \partial x_2} = \frac{\partial^2 P_4}{\partial v_1 \partial v_2} \frac{\partial v_1}{\partial z_1} \frac{\partial v_2}{\partial x_2} - \sum_{\mathcal{S} \subset \mathcal{J}_1, \mathcal{S} \neq \emptyset} \frac{\partial^2 P_{5,1}^{\mathcal{S}}}{\partial v_1 \partial v_2} \frac{\partial v_1}{\partial z_1} \frac{\partial v_2}{\partial x_2} \tag{28}$$

Combining (27) and (28), we obtain

$$\frac{\partial^2 s_1}{\partial z_1 \partial z_2} \Big/ \frac{\partial^2 s_1}{\partial z_1 \partial x_2} = \frac{\frac{\partial v_2}{\partial z_2}}{\frac{\partial v_2}{\partial x_2}} \tag{29}$$

Since this equation holds for all $(\mathbf{x}, \mathbf{z})$ 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 $k = 1, \cdots, J$ and $x_j = 0$ for all $j \neq 2$ in (26), we can write

$$\frac{\partial s_1}{\partial z_1} = f(g(x_2)) \tag{30}$$

where

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

and $g(x_2) : x_2 \mapsto v(x_2, 0)$. 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 f with coefficients depending on the derivatives of g and derivatives of f of order strictly less than $n-1$. Further, by the above, all derivatives of g are known. Thus, we have a system of equations that can be uniquely solved for the derivatives of f by recursion.

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

$$\frac{\partial s_1}{\partial z_1} = f(v(x_2, z_2)) \tag{31}$$

where now note that z_2 is no longer fixed at 0. Again by the chain rule we have that, for $n \geq 3$, $\frac{\partial^n s_1}{\partial z_1 \partial z_2 \partial x_2^{n-2}}$ is a linear function of $\frac{\partial^{n-1} v}{\partial z_2 \partial x_2^{n-2}}(0,0)$ with coefficients depending on lower-order derivatives of v as well as derivatives of f. Because all derivatives of f 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 $l \geq 2$, by differentiating (31) l times wrt z_2 and again $n-l-1$ times wrt x_2, one can write $\frac{\partial^n s_1}{\partial z_1 \partial z_2^l \partial x_2^{n-l-1}}$ as a linear function of $\frac{\partial^{n-1} v}{\partial z_2^l \partial x_2^{n-l-1}}(0,0)$ with known coefficients. This system can then be solved iteratively for $\frac{\partial^{n-1} v}{\partial z_2^l \partial x_2^{n-l-1}}(0,0)$ for all $n \geq 4, l \geq 2$.

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

A.2 Proof of Theorem 3

Note that the analog of equation (25) in the random coefficients model holds for any given value of $\alpha > 0$, $\beta > 0$, so that we can write:

$$\begin{aligned} s_1 = \int \Bigg[& P(U_1 \geq U_k \ \forall k) - \sum_{\mathcal{S} \subset \mathcal{J}_1, \mathcal{S} \neq \emptyset} P(\{U_1 \geq U_k \forall k\} \cap \{VU_j \geq VU_1 \text{ for at least one } j \in \mathcal{S}\} \\ & \cap \{g_i(x_1, \epsilon_1, U_j) \leq 0 \text{ for all } j \in \mathcal{S}\} \cap \{g_i(x_1, \epsilon_1, U_j) \geq 0 \text{ for all } j \in \mathcal{J}_1 \backslash \mathcal{S}\}) \Bigg] dF_{\alpha,\beta}(\alpha, \beta) \\ \equiv & \int \left[P_6(\mathbf{v}; \alpha, \beta) - \sum_{\mathcal{S} \subset \mathcal{J}_1, \mathcal{S} \neq \emptyset} P_7^{\mathcal{S}}(\mathbf{v}, x_1; \alpha, \beta) \right] dF_{\alpha,\beta}(\alpha, \beta) \end{aligned} \tag{32}$$

where $\mathbf{v} \equiv (v_1, \cdots, v_J)$ and in the model considered here we have $v_j \equiv x_j\alpha_i + z_j\beta_i$. Thus, for all integers $n \geq \tilde{n} \geq 0$, we have:

$$\frac{\partial^{1+n} s_1}{\partial z_1 \partial z_2^{\tilde{n}} \partial x_2^{n-\tilde{n}}} = \int \frac{\partial^{1+n} \left[P_6(\mathbf{v}; \alpha, \beta) - \sum_{\mathcal{S} \subset \mathcal{J}_1, \mathcal{S} \neq \emptyset} P_7^{\mathcal{S}}(\mathbf{v}, x_1; \alpha, \beta) \right]}{\partial z_1 \partial z_2^{\tilde{n}} \partial x_2^{n-\tilde{n}}} dF(\alpha, \beta) \tag{33}$$

Then, letting $h(\mathbf{v}) \equiv P(v_{i1} + \epsilon_{i1} \geq v_{ik} + \epsilon_{ik} \; \forall k)$ and $K \equiv K_\alpha K_\beta$, we can evaluate (33) at $(\mathbf{x}, \mathbf{z}) = (\mathbf{0}, \mathbf{0})$ and re-write it as

$$\begin{aligned} \frac{\partial^{1+n} s_1}{\partial z_1 \partial z_2^{\tilde{n}} \partial x_2^{n-\tilde{n}}} &= \frac{\partial^{1+n} h}{\partial v_{i1} \partial v_{i2}^{n}}(\mathbf{0}) \sum_{k_\alpha=1}^{K_\alpha} \sum_{k_\beta=1}^{K_\beta} \alpha_{k_\alpha}^{n-\tilde{n}} \beta_{k_\beta}^{\tilde{n}+1} \tilde{\pi}_{k_\alpha, k_\beta} + \sum_{k_\alpha=1}^{K_\alpha} \sum_{k_\beta=1}^{K_\beta} \alpha_{k_\alpha}^{n-\tilde{n}} \beta_{k_\beta}^{\tilde{n}+1} \tilde{\pi}_{k_\alpha, k_\beta} \sum_{\mathcal{S} \subset \mathcal{J}_1, \mathcal{S} \neq \emptyset} \frac{\partial^{1+n} P_7^{\mathcal{S}}(\mathbf{0}, 0; \alpha_{k_\alpha}, \beta_{k_\beta})}{\partial v_{i1} \partial v_{i2}^{n}} \\ &= \sum_{k=1}^{K} \left[a_{k,n,\tilde{n}} + b_{n,\tilde{n}} f_{k,n} \right] \pi_k \end{aligned} \tag{34}$$

for known $K-$vectors $a_{n,\tilde{n}}, b_{n,\tilde{n}}$ and unknown $K-$vectors f_n, π.

Setting $n = K - 1$ and stacking the equations corresponding to $\tilde{n} = 0, \cdots, K-1$,[19] we get

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

where q is a known $K-$vector, A, B are known $K-$by$-K$ matrices, and $f * \pi$ denotes the element-by-element product of $f = [f_{1,K-1}, \cdots, f_{K,K-1}]'$ and π. I re-write this system of equations in a way that highlights which objects depend on $z \equiv (z_1, \cdots, z_J)$ as follows

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

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

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

Combining the last two systems, we get

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

If B is full rank,[20] we obtain identification of $(f(\mathbf{z} = \mathbf{\Delta z}) - f(\mathbf{z} = \mathbf{0})) * \pi$. Also, note that, for all $k \in$

[19] The symbol $\frac{\partial}{\partial q^0}$ means that no derivative is taken wrt q.

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

$\{1, \cdots, K\}$, letting $z_{j:j' \equiv (z_j, \cdots, z_{j'})}$

$$\begin{aligned}
& \lim_{\Delta z \to 0} \frac{f_k(\mathbf{z} = \mathbf{\Delta z}) - f_k(\mathbf{z} = \mathbf{0})}{\Delta z} \\
= & \sum_{t=1}^{J} \lim_{\Delta z \to 0} \frac{f_k(z_{1:t} = \mathbf{\Delta z}, \mathbf{z}_{t+1:J} = \mathbf{0}) - f_k(z_{1:t-1} = \mathbf{\Delta z}, \mathbf{z}_{t:J} = \mathbf{0})}{\Delta z} \\
= & \sum_{t=1}^{J} \frac{\partial f_k}{\partial z_t} (\mathbf{z} = \mathbf{0})
\end{aligned} \tag{35}$$

where the second equality follows from the Moore-Osgood Theorem under the assumption that the ratios inside the sum on the second line converge uniformly to $\frac{\partial f_k}{\partial z_t}(z)$. Thus, we can write

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

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

Next, by taking another derivative wrt z_1 in (34), we can write

$$q_{(1)}(\mathbf{z} = \mathbf{0}) = A_{(1)}\pi + B_{(1)} \left(\frac{\partial f}{\partial z_1} (\mathbf{z} = \mathbf{0}) * \pi \right) \tag{36}$$

for known $K-$by$-K$ matrices $A_{(1)}, B_{(1)}$ and a known $K-$vector $q_{(1)}(\mathbf{z} = \mathbf{0})$. Similarly, taking another derivative wrt z_j for $j \in \mathcal{J}_1$ in (34), we can write

$$q_{(j)}(\mathbf{z} = \mathbf{0}) = A_{(j)}\pi + B_{(j)} \left(\frac{\partial f}{\partial z_j} (\mathbf{z} = \mathbf{0}) * \pi \right) \tag{37}$$

Note that $B_{(j)} = B$ for all $j \in \mathcal{J}$ and so we can write

$$\sum_{j=1}^{J} q_{(j)}(\mathbf{z} = \mathbf{0}) = \left(\sum_{j=1}^{J} A_{(j)} \right) \pi + B \left[\left(\sum_{j=1}^{J} \frac{\partial f_k}{\partial z_j} (\mathbf{z} = \mathbf{0}) \right) * \pi \right] \tag{38}$$

From above, $\left(\sum_{t=1}^{J} \frac{\partial f_k}{\partial z_t} (\mathbf{z} = \mathbf{0}) \right) * \pi$ is identified. This implies that π is identified if the matrix $\sum_{j=1}^{J} A_{(j)}$ is invertible.[21]

A.3 Recovery of Search Costs Given Preferences in the Weitzman Model

Suppose for example 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),[22] the optimal search strategy is for consumers to behave as if they were

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

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

choosing among options in a static model with utilities given by $\tilde{U}_{ij} = x_j\alpha + \min\{z_j, r_i\}\beta + \epsilon_{ij}$, where r_i denotes i's reservation value in units of z (see Example 1). Thus, dropping i subscripts, ordering goods so that $z_1 \geq z_2 \geq \cdots \geq z_J$, and letting

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

we can write

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

where F_r denotes the cdf of r and the second equality assumes that search costs (and thus r) are independent of ϵ. Therefore, we have

$$\frac{\partial s_1}{\partial z_1} = \left[\frac{\partial}{\partial z_1} P(\epsilon_k - \epsilon_1 \leq (x_1 - x_k)\alpha + (z_1 - z_k)\beta \ \forall k)\right] P(r \geq z_1) \tag{39}$$

Given identification of (α, β) by the argument in Section 2.1, the first term on the rhs of (39) is identified given parametric assumptions on the distribution of ϵ. Thus, $P\{r \geq z_1\}$ is identified. Repeating the argument for all z_1, one can trace out the entire distribution of r. Since c, the search cost for consumer i, is a known transformation of r,[23] the distribution of c is also identified.

Equation (39) 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 (39) is invariant to increasing all z_j's by the same amount. Thus, we can write

$$\frac{\frac{\partial s_1}{\partial z_1}(\mathbf{z} + \mathbf{\Delta})}{\frac{\partial s_1}{\partial z_1}(\mathbf{z})} = \frac{P(r \geq z_1 + \Delta)}{P(r \geq z_1)} \tag{40}$$

where $\mathbf{\Delta}$ is a $J-$vector with all elements equal to some Δ. Letting $\Delta \to -\infty$, the numerator on the rhs of (40) goes to 1, which yields identification of $P(r \geq z_1)$. Repeating the argument for all z_1, one can trace out the entire distribution of r and recover the distribution of c as above.

A.4 Proof of Theorem 1 when Observables Impact Search but not Utility

The proof proceeds exactly as in Section A.1 except for the steps given below. Here, we demonstrate that we can decompose s_1 into a component which depends on x_j and z_j only via U_j and a component which

[23] 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.

does not depend on z_1.

As in Section 3.1, suppose we have at least 2 goods, and there is at least one good $j \neq 1$ such that $r_j > r_1$ so that the set R of goods with $r_j > r_1$ is non-empty. 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 (3) we can write for all $(\mathbf{x}, \mathbf{z})$ with $z_1 \geq z_j$ for all j:

$$\begin{aligned} s_1 = P(U_1 \geq U_k \ \forall k \in R) - \sum_{\mathcal{S} \subset \mathcal{J}_1, \mathcal{S} \neq \emptyset} & P\left(\{U_1 \geq U_k \forall k \in R\} \cap \{m(VU_j, r_j) \geq m(VU_1, r_1) \text{ for at least one } j \in \mathcal{S}\}\right. \\ & \left. \cap \{g_i(x_1, \epsilon_1, U_j) \leq 0 \text{ for all } j \in \mathcal{S}\} \cap \{g_i(x_1, \epsilon_1, U_j) \geq 0 \text{ for all } j \in \mathcal{J}_1 \backslash \mathcal{S}\}\right) \\ \equiv \ & P_{4new}(\mathbf{v}) - \sum_{\mathcal{S} \subset \mathcal{J}_1, \mathcal{S} \neq \emptyset} P_{5new}^{\mathcal{S}}(\mathbf{v}, x_1) \end{aligned} \tag{41}$$

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

$$\begin{aligned} P_{5new}^{\mathcal{S}} = \ & P\left(\{U_1 \geq U_k \ \forall k \in R\} \cap \{g_i(x_1, \epsilon_1, U_j) \leq 0 \text{ for all } j \in \mathcal{S}\} \cap \{g_i(x_1, \epsilon_1, U_j) \geq 0 \text{ for all } j \in \mathcal{J}_1 \backslash \mathcal{S}\}\right) - \\ & P\left(\{U_1 \geq U_k \ \forall k \in R\} \cap \{m(VU_1, r_1) \geq m(VU_j, r_j) \text{ for all } j \in \mathcal{S}\}\right. \\ & \left. \cap \{g_i(x_1, \epsilon_1, U_j) \leq 0 \text{ for all } j \in \mathcal{S}\} \cap \{g_i(x_1, \epsilon_1, U_j) \geq 0 \text{ for all } j \in \mathcal{J}_1 \backslash \mathcal{S}\}\right) \\ = \ & P\left(\{U_1 \geq U_k \ \forall k \in R\} \cap \{g_i(x_1, \epsilon_1, U_j) \leq 0 \text{ for all } j \in \mathcal{S}\} \cap \{g_i(x_1, \epsilon_1, U_j) \geq 0 \text{ for all } j \in \mathcal{J}_1 \backslash \mathcal{S}\}\right) - \\ & P\left(\{m(VU_1, r_1) \geq m(VU_j, r_j) \text{ for all } j \in \mathcal{S}\} \cap \{g_i(x_1, \epsilon_1, U_j) \leq 0 \text{ for all } j \in \mathcal{S}\} \cap \{g_i(x_1, \epsilon_1, U_j) \geq 0 \text{ for all } j \in \mathcal{J}_1 \backslash \mathcal{S}\}\right) \\ \equiv \ & P_{5new,1}^{\mathcal{S}}(\mathbf{v}, x_1) - P_{5new,2}^{\mathcal{S}}(\mathbf{v}_{-1}, x_1) \end{aligned}$$

where $\mathbf{v}_{-1} \equiv \mathbf{v} \setminus v_1$. This argument exactly parallels the argument in Section A.1, except now we have additionally used the fact that for $j \in \mathcal{S}$, if $j \in R$, then $m(VU_1, r_1) \geq m(VU_j, r_j)$ implies $VU_1 \geq VU_j$, which in turn implies $U_1 \geq U_j$. Note that $P_{5new,2}^{\mathcal{S}}$ does not depend on z_1 and $P_{5new,1}^{\mathcal{S}}(\mathbf{v}, x_1)$ only depends on x_j via U_j, so the remainder of the argument in Section A.1 applies.

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

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. Lockwood, and D. Taubinsky (2019). Regressive sin taxes, with an application to the optimal soda tax. Technical report, National Bureau of Economic Research.

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.

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.

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.

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.

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 (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., 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. *The American Economic Review 105*(8), 2449–2500.

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.

Honka, E. and P. Chintagunta (2016). Simultaneous or sequential? search strategies in the us auto insurance industry. *Marketing Science 36*(1), 21–42.

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.

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.

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.

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.

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.