NBER WORKING PAPER SERIES

PARTIAL IDENTIFICATION OF HETEROGENEITY IN PREFERENCE ORDERINGS OVER DISCRETE CHOICES

Itai Sher
Jeremy T. Fox
Kyoo il Kim
Patrick Bajari

Working Paper 17346
http://www.nber.org/papers/w17346

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2011

Fox thanks the National Science Foundation for generous funding. Bajari thanks the National Science Foundation, grant SES-0720463, for generous research support. We are grateful to Joerg Stoye and Rakesh Vohra for useful comments related to the issues in this paper. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Partial Identification of Heterogeneity in Preference Orderings Over Discrete Choices
Itai Sher, Jeremy T. Fox, Kyoo il Kim, and Patrick Bajari
NBER Working Paper No. 17346
August 2011
JEL No. C25,L0

ABSTRACT

We study a variant of a random utility model that takes a probability distribution over preference relations as its primitive. We do not model products using a space of observed characteristics. The distribution of preferences is only partially identified using cross-sectional data on varying budget sets. Imposing monotonicity in product characteristics does not restore full identification. Using a linear programming approach to partial identification, we show how to obtain bounds on probabilities of any ordering relation. We also do constructively point identify the proportion of consumers who prefer one budget set over one or two others. This result is useful for welfare. Panel data and special regressors are two ways to gain full point identification.

Itai Sher
Department of Economics
University of Minnesota
4-145 Hanson Hall
1925 Fourth Street South
Minneapolis, MN 55455
isher@umn.edu

Jeremy T. Fox
Economics Department
University of Michigan
238 Lorch Hall
611 Tappan St.
Ann Arbor, MI 48104
and NBER
jeremyfox@gmail.com

Kyoo il Kim
Department of Economics
University of Minnesota
4-129 Hanson Hall
1925 4th Street South
Minneapolis, MN 55455
kyookim@umn.edu

Patrick Bajari
Professor of Economics
University of Minnesota
4-101 Hanson Hall
1925 4th Street South
Minneapolis, MN 55455
and NBER
bajari@econ.umn.edu

1 Introduction

Researchers in many fields estimate random utility models of multinomial choice (McFadden 1972). These models typically relate a discrete choice out of a finite number J of alternatives to a vector of H product characteristics. Modeling utility as being a heterogeneous function over a product space X has two main advantages. First, preferences for a consumer can be captured by a parsimonious set of preferences over the vectors $x_j \in X$, which aids estimation when a limited number of observations is available. Second, the estimated model can be used to compute the demand for new goods, meaning combinations of product characteristics x_j not observed in the prior data.

Modeling products in characteristics space is empirically useful as long as the researcher has data on all of the x_j (or all but one) relevant to consumers.[1] However, in many applications the researcher finds it difficult to measure the appropriate x_j. In the autos applications of Berry, Levinsohn and Pakes (1995) and Petrin (2002), the data measure car characteristics such as fuel economy and speed but do not measure hard-to-quantify but potentially important car characteristics such as physical attractiveness, quality, luxurious feel, and handling. In the breakfast cereal application of Nevo (2001), the author measures characteristics such as sugar content, fat content, advertising and mushiness. However, as Hitsch (2006) models, the appeal of a breakfast cereal arises from intangible characteristics that manufacturers may have little information on before product launch.

Goettler (1999) considers the problem of consumers faced with a discrete choice of television programs. It is not immediately clear how to specify the characteristics of *Seinfeld*, *Friends*, *CSI*, *Law and Order*, *The O'Reilley Factor* or *Face The Nation*. Some characteristics are obvious, such as whether the show is comedy, drama or news. However, how should one specify the characteristics of alternative comedies such as *Seinfeld* versus *Friends*? Is it appropriate to think of both *The O'Reilley Factor* and *Face The Nation* as "news" when *Face The Nation* is primarily based on interviews with leading opinion makers while *The O'Reilley Factor* is usually based on interviews with pundits? Coming up with a finite and exhaustive list of characteristics is often difficult.

Motivated by the difficulty of measuring all product characteristics, we work with an alternative representation of random utility models that has been proposed in mathematical economics and mathematical psychology (Block and Marschak 1960, Manski 1977, Falmagne 1978, Barbera and Pattanaik 1986, McFadden and Richter 1991, Fishburn 1998, McFadden 2005, Manski 2007). In these papers, the primitive used for modeling consumer i's choice behavior is a strict preference ordering $\succ_i$ on alternatives. It is well-known that a preference ordering $\succ_i$ induces a utility function u_i (up to strictly monotone transformations). Consumer i chooses good j if $j \succ_i j'$ for all $j' \neq j$; this corresponds to the situation where $u_{i,j} > u_{i,j'}$ in a model where preferences are represented by a utility function. In this literature, it is common to restrict attention to strict orderings, which rule out indifference between distinct alternatives. The primitive parameter of the model is $p(\succ_i)$, a probability distribution on

[1] For example, the approach of Berry, Levinsohn and Pakes (1995) allows for H characteristics in x_j and a single, unobserved characteristic ξ_j.

preference orderings $\succ_i$, which plays the role of $F_u(u_{i,1}, ..., u_{i,J})$, a probability distribution on utilities, in a random utility model. Simple combinatorics implies that if there are J choices, then there are $J!$ ways to rank these choices. Hence, the primitives of our model will be the $J!$ probabilities, $p(\succ_i)$, induced by different types of consumers. Most of our results do not consider product characteristics x_j, as the previous empirical literature does. This includes not focusing on prices for our main results, as prices are not used or do not flexibly vary in some markets, such as the television market studied by Goettler (1999).

As a simple example, suppose that a consumer has the choice between three types of cars {GM, Ford, Chrysler}. Consumer i ranks the choices as Ford$\succ_i$GM$\succ_i$Chrysler while consumer i' ranks the choices as Ford$\succ_{i'}$Chrysler$\succ_{i'}$GM. With three available choices, there are $3! = 6$ possible ways to rank these choices. The parameters of the model are a set of six probabilities (or five independent probabilities as they sum to 1), which we denote as $p(\succ)$. Given the model parameters, $p(\succ)$, and a set of choices, we can compute choice probabilities or aggregate demand in a straightforward manner. The probability that Ford is chosen is equal to the probability of the ranking Ford$\succ$GM$\succ$Chrysler plus the probability of the ranking Ford$\succ$Chrysler$\succ$GM.

As mentioned above, it is well known that every vector of utilities $(u_{i,1}, ..., u_{i,J})$ is associated with a unique preference ordering $\succ_i$. Conversely, when the choice set is finite, every $\succ_i$ can be associated with (one of many) vector of utilities $(u_{i,1}, ..., u_{i,J})$ that will generate the same choice behavior. Therefore, passing between preference orderings and utility values amounts to a change of variables. However, working with preference orderings is more convenient for two reasons. First, the analysis of the problem of identification attains greater clarity. As is well known from classical consumer theory (e.g. Mas-Colell, Whinston and Green 1995), it is always possible to make a positive monotonic transformation of utility and leave choice behavior unaltered. Differences in utility that do not manifest themselves in differences in preference rank are not behaviorally meaningful, and cannot be identified from demand. It follows that in a random utility model, the distribution of random utilities $F_u(u_{i,1}, ..., u_{i,J})$ cannot be identified from choice behavior, unless we restrict attention to a set of utility functions such that no pair is related by a monotonic transformation (Matzkin 1993). For this reason, identification in discrete choice models is simplified and clarified by working with $p(\succ_i)$ instead of $F_u(u_{i,1}, ..., u_{i,J})$. Second, we shall show that there is a natural linear relationship between consumer demand and $p(\succ_i)$. This linearity simplifies the analysis of identification, and would also simplify the estimation of properties of the distribution that are point or partially identified.

Our paper studies the identification of $p(\succ_i)$, the distribution of preferences. While the work in mathematical economics mentioned above studies the same model, the results from this literature are not sufficiently integrated into the modern empirical literature. We both review results on identification from this prior literature (especially Falmagne 1978) and prove new identification results about p. Our focus on the identification of $p(\succ_i)$ contrasts with much of the earlier work, which focuses on the identification of choice probabilities under new budget sets (Manski 2007) or, especially, on testing whether there exists a $p(\succ_i)$ consistent with the data (e.g., Fishburn 1992, McFadden 2005).

Following the earlier literature on preference orderings, we primarily consider the case that the economist observes a cross section of choice probabilities from various budget sets, meaning at most every nonempty subset of the finite set of the J alternatives. While it is common in the economic theory literature to assume an analyst has data on all budget sets, it is less common to have this in empirical work. One setting where rich budget set variation is observed is experimental data. Budget sets are often varied in economic experiments, whether the experiment is for academic research or for commercial analysis of demand. Budget set variation is also found in television viewing, where a choice situation is a time period and different shows compete for viewers in each time period.

We establish that for four or more choices, data on choice probabilities for all budget sets are not sufficiently rich to identify the population distribution over preferences, $p(\succ)$. While the prior literature in mathematical economics was generally aware of underidentification issues, we formally explore the degree of underidentification. This result implies that for more than three choices, the identification of random utility models hinges crucially on restricting the model to focus on preferences (up to a functional form) over observed product characteristics x_j. As this result is a non-identification result using a large amount of budget set variation, it suggests that existing empirical work makes strong use of observed product characteristics or functional form assumptions on utility.

Having established that $p(\succ)$ is rarely point identified, we explore its identifiable features. We begin by reviewing some results of Flamagne (1978) that point identify features of p. In particular, these results establish that (i) the probabilities of upper and lower contour sets are identified, and (ii) the marginal distribution of the rank ordering of each product is identified; that is one can identify the probability that choice j is the third most popular choice, for example. Note that the main focus of the prior literature to which these results belong was testing rationality rather than identification.[2] However, these results provide background and foundation for our original results.

Our paper is part of a partial identification approach in the literature (see e.g. Manski 1995, 2003). The most related paper to ours in this approach is Manski (2007). He explores bounds on choice probabilities under counterfactual choice situations while we focus on identification of the primitive, probability distribution of preference. We further extend a linear programming approach to partial identification in this setting introduced by Manski (2007). We show how to find the maximum and minimum probabilities of any ordering relation consistent with the data using linear programs. Similarly, we show how to find the extremal probabilities, again consistent with the data, for various properties of preference both under the assumption that the researcher observes the distribution of choices from all budget sets and under the assumption that the researcher observes the distribution of choices from only a subset of budget sets. The linear programs used for this purpose can be expressed either in terms of choice probabilities or in terms of lower and upper contour set probabilities.

[2] Manski (1977) works with the primitive of a measure over pairs of budget sets and decision makers. He does not explore identification of this measure. Even establishing identification of this object would not establish identification of p, as a choice set with n elements corresponds to $n!$ different preference orderings.

The failure to attain full point identification does not necessarily prevent all welfare comparisons. Given any three budget sets, we can point identify the distribution of consumer preferences over these budget sets. In other words, we can identify the "voting preferences" of consumers for any three budget sets. We provide simple explicit expressions for these distributions. This is an important identification result for welfare analysis, as many policies may alter the choice sets available to consumers. Our result does not require observed product characteristics or prices.

One possible solution to the lack of point identification of $p(\succ)$ might be to return to using data on vectors of H product characteristics x_j for each choice. Rather than assuming some functional form for the utility of j as a function of x_j, one might be willing to impose the assumption that utility is monotone in the H elements of x_j, for each consumer. We formally investigate monotonicity restrictions and show, regrettably, that they do not lead to the point of identification of $p(\succ)$. If the true data generating process has a subset A of more than three elements such that no element of A dominates any other element of A–in the sense of being weakly better along all characteristics–then $p(\succ)$ cannot be identified.

We relate our results to the prior empirical literature by emphasizing two stronger forms of data variation and modeling assumptions that lead to point identification of the distribution of preferences. One set of identifying data and model assumptions is that utility is quasi-linear in prices and that prices flexibly vary across all choice situations. A second set of assumptions allows for panel data: the same consumer is observed making choices at different budget sets.

Unlike most of the previous literature in mathematical economics on random preferences, our focus is on exploring the identification of $p(\succ)$. Thus, in broad topic our paper relates to results on nonparametric identification in multinomial choice models where observed product characteristics x_j provide a major source of identifying power. This literature on observed product (or consumer) characteristics discusses identifying the distribution of utility functions or random coefficients on x_j (Ichimura and Thompson 1998, Matzkin 2007, Bajari, Fox, Kim and Ryan 2010, Fox and Gandhi 2010, Chiappori and Komunjer 2009) or of utility values conditional on x_j (Lewbel 2000, Berry and Haile 2010).

2 Model

Consider a household i who chooses between a finite and mutually exclusive set of alternatives $j = 1, ..., J$. Let $\mathcal{J} = \{1, ..., J\}$ denote the set of alternatives. For example, the choice might be between alternative computer systems such as a Dell, HP or Apple. In most cases, it is natural to assume that the choice of the household is constrained because not all products are available in all markets. Let $B \subseteq \mathcal{J}$ denote households' choice set. Throughout, we assume that all households within a given market face the same choice set.

Each consumer has a strict preference relation $\succ$ over $\mathcal{J}$. For the J choices, there are $J!$ ways to uniquely order the available choices from highest to lowest. As discussed earlier, preference orderings can be represented by utility functions. In a discrete choice setting, u orders the choices: $j \succ k$ if and only if $u_j > u_k$. We allow consumers to have any strict preference ordering. For instance, one group of consumers may order the set of personal computers as Dell$\succ$HP$\succ$Apple while others may have the order Apple$\succ$HP$\succ$Dell.

Let $\mathcal{R}$ be the set containing all of the $J!$ linear orderings on $\mathcal{J}$. Let $p : \mathcal{R} \to [0,1]$ be a mass function where $p(\succ)$ is the fraction of the consumers that have preferences corresponding to the ordering $\succ$, where $\sum_{\succ \in \mathcal{R}} p(\succ) = 1$. For any subset L of $\mathcal{R}$, define $p(L) = \sum_{\succ \in L} p(\succ)$.

Let $\Pr(j \mid B, p)$ be the population probability of j being chosen given a choice set B and a measure p over the orderings. For each $j \in B$, we have:

$$\Pr(j \mid B, p) = \sum_{\succ \in \mathcal{R}} p(\succ) 1 \{ j \succ j' \text{ for all } j' \in B, j' \neq j \} . \quad (1)$$

In (1), $1 \{ j \succ j' \text{ for all } j' \in B, j' \neq j \}$ is the indicator function for the event that j is the maximizing alternative in B. Thus the probability that j is the demanded alternative in B is equal to the sum of probabilities of drawing elements of $\mathcal{R}$ that make j a maximizer.

We could generalize the above model by allowing $p(\succ)$ to be conditional on observed covariates s such as demographics or advertising levels that might influence consumer preferences. For instance, we might expect families with young children to have a stronger preference for minivans than single men (Petrin 2002). We shall suppress these arguments, although our approach could be seen as seeking to learn the distribution of preferences for minivans within, say, the class of families with young children.

3 Identification

This section asks whether aggregated consumer choice behavior is sufficient to uniquely recover p, the distribution of the individual heterogeneity. We explore a situation where the same (finite or infinite) population of consumers is present in every choice situation, but there are many choice situations where the budget sets are varied so that (at most) every non-empty $B \subseteq \mathcal{J}$ is offered to consumers. As discussed in the introduction, this is the data scheme in the previous literature on this model. In each choice situation or market indexed by B, we observe the choice probabilities or market shares of each choice $j \in B$. This also corresponds to the case where we have individual data but cannot track consumers across choice situations using a panel identifier. As discussed in the introduction, variation in choice sets B is common in experimental applications and in applications such as television viewing.

Definition 1.

- We say that the measure p is **identified from choice probability data** if $p \neq p'$ implies that $\Pr(j \mid B, p) \neq \Pr(j \mid B, p')$ for some B and j.

- We say that the probability of the set O of orderings is **identified from choice probability data** if $p(O) \neq p'(O)$ implies that $\Pr(j \mid B, p) \neq \Pr(j \mid B, p')$ for some B and j.

We will show that the measure p is not generally point identified if there are four or more alternatives. However, we will also show that important features of the distribution of consumer preferences are identified.

3.1 Identification of Lower Contour Set Probabilities

In this section, we review previous results by Block and Marschak (1960), Falmagne (1978), Barbera and Pattanaik (1986) that bear on the partial identification of the distribution of preference orderings from observations on the distribution of choices from budget sets. These results provide background and foundation for the results that we provide in later sections.

For any preference order $\succ$ and item $j \in \mathcal{J}$, define the **strict lower contour set of** j **given** $\succ$ as

$$L(j, \succ) := \{i \in \mathcal{J} : j \succ i\}.$$

Then define

$$\begin{aligned} C_{j,K} &:= \{\succ \in \mathcal{R} : K \subseteq L(j, \succ)\}, \\ L_{j,K} &:= \{\succ \in \mathcal{R} : K = L(j, \succ)\} \qquad \forall K \subseteq \mathcal{J}, \forall j \notin K. \end{aligned}$$

$C_{j,K}$ is the event that K is contained in the lower contour set of j, or in other words, the event that j would be chosen over all elements of K. The probability of the events $C_{j,K}$ is immediately identified from the data. Indeed, we have

$$\Pr(j|B, p) = p(C_{j,B\setminus j}). \tag{2}$$

$L_{j,K}$ is the event that the lower contour set of j is *exactly* K. Note that the events $L_{j,K}$ are finer than the events $C_{j,K}$, whose probabilities are directly observable from choice. More precisely, $L_{j,K} \subseteq C_{j,K}$, and the inclusion is strict except when $K = \mathcal{J} \setminus j$. The following example illustrates the above concepts.

Example 2. Let $\mathcal{J} = \{1, 2, 3, 4, 5\}$. Suppose we observe that from budget set $\{1, 3, 5\}$, 1/3 of the population chooses alternative 1. This tells us that 1/3 of the population prefers 1 to everything in the set $\{3, 5\}$, or in other words that $\Pr(1|\{1, 3, 5\}, p) = p(C_{1,\{3,5\}}) = 1/3$. However, the above observation alone does not reveal the proportion of the population that prefers 1 to the items in $\{3, 5\}$, but does not prefer 1 to any other alternative; in other words, the above observation does not reveal $p(L_{1,\{3,5\}})$, the probability that $\{3, 5\}$ is the lower contour set of 1. All we can infer is that $p(L_{1,\{3,5\}}) \leq 1/3$. It is possible that $p(L_{1,\{3,5\}}) = 1/3$ if everyone who prefers 1 to 3 and 5 also prefers 2 and 4 to 1. It is also possible that $p(L_{1,\{3,5\}}) = 0$, if, for example, everyone who prefers 1 to 3 and 5 also prefers 1 to 2. Below, we will show that we can identify $p(L_{1,\{3,5\}})$.

The following identification result follows from results of Block and Marschak (1960) and Falmagne (1978). (See in particular, Definition 2 and Theorem 2 of Flamagne 1978).

Theorem 3. *1. The probabilities $p(L_{j,K})$ are identified via the equation:*

$$p(L_{j,K}) = \sum_{S \subseteq \mathcal{J} \setminus (K \cup j)} (-1)^{|S|} \Pr(j|K \cup S \cup j, p) \qquad \forall K \subseteq \mathcal{J}, \forall j \notin K. \tag{3}$$

2. Conversely, choice probabilities are determined by the probabilities $p(L_{j,K})$ via:

$$\Pr(j|B, p) = \sum_{\substack{K \supseteq B \setminus j: \\ K \not\ni j}} p(L_{j,K}), \qquad \forall j \in \mathcal{J}, \forall B \ni j. \tag{4}$$

Although Theorem 3 is not new, the appendix contains a proof of the theorem for completeness. In (3), $|S|$ is the number of elements in S, so $\Pr(j|K \cup S \cup j, p)$ enters positively or negatively into the expression for $p(L_{j,K})$ depending on whether S has an even or odd number of elements. Note that $S = \emptyset$ is included in the summation. Part 1 of the theorem show that it is possible to identify the probability events of the form $L_{j,K}$ which are finer than those of the form $C_{j,K}$ for which we directly observe probabilities in the data. Part 2 of the theorem shows that if we know the probability of all lower contour sets, we can predict all choice probabilities. So, parts 1 and 2 of the theorem together show that this identification result is *sharp* in the sense that knowing the probabilities of all lower contour sets is equivalent to knowing all choice probabilities. Note finally that in order to identify $p(L_{j,K})$, it is sufficient to observe $\Pr(j|B, p)$ for $B \supseteq K \cup j$. In other words, it is sufficient to observe choice from budget sets B containing $K \cup j$, and moreover, one needs only to observe the probability that j is chosen from such budget sets B and not the probability that any other item is chosen from B.

The expressions on the right-hand side of (3) are variants of expressions known as **Block-Marschak polynomials**. Whereas the expressions above are defined in terms of lower contour sets, the Block-Marschak polynomials were originally defined using analogous expressions in terms of upper rather than lower contour sets. Upper contour set probabilities are identified in a manner analogous to lower contour set probabilities.[3]

In concluding this section we mention a corollary of Theorem 3 which is also due to Flamagne (1978) (see Theorem 7 of that paper) and which should be useful for applied work. Let $r(j, \succ)$ give

[3] In particular upper contour sets are defined as follows:

$$U(j, \succ) := \{i \in \mathcal{J} : i \succ j\}.$$
$$U_{j,K} := \{\succ \in \mathcal{R} : K = U(j, \succ)\} \qquad \forall K \subseteq \mathcal{J}, \forall j \notin K.$$

the ranking of choice j in the preference ordering $\succ$, with the most preferred item being ranked 1 and the least preferred being ranked J. Let

$$\Pr\left(r\left(j,\succ\right)=k\right)=\sum_{\succ\in\mathcal{R}}p\left(\succ\right)1\left\{r\left(j,\succ\right)=k\right\}$$

be the probability that a consumer chosen at random has choice j in the kth ranking slot.

Corollary 4. *The quantity* $\Pr\left(r\left(j,\succ\right)=k\right)$ *is identified from data on choice probabilities.*

This is a consequence of Theorem 3 because:

$$\Pr\left(r\left(j,\succ\right)=k\right)=\sum_{K:|K|=J-k}p(L_{j,K}).$$

3.2 Partial Identification: A Linear Programming Approach

In this section we sketch a linear programing approach to partial identification in our setting. Let $\mathcal{S}$ be any family of subsets of $\mathcal{R}$ such that

$$\mathcal{R}\in\mathcal{S}. \tag{5}$$

For any $S\in\mathcal{S}$, let $p(S)$ be the probability that $\succ\in S$. Then consider the linear system:

$$\sum_{\succ\in S}\pi_{\succ}=p(S) \qquad \forall S\in\mathcal{S} \tag{6}$$

$$\pi_{\succ}\geq 0 \qquad \forall\succ\in\mathcal{R}. \tag{7}$$

Interpreting $\pi_{\succ}$ as the probability that a consumer has preference ordering $\succ$, any feasible solution $\pi=(\pi_{\succ}:\succ\in\mathcal{R})$ to (6-7) corresponds to a probability distribution that assigns probability $p(S)$ to each set $S\in\mathcal{S}$. Notice in particular that (5) and $p(\mathcal{R})=1$ imply that (6) includes the equation:

$$\sum_{\succ\in\mathcal{R}}\pi_{\succ}=1, \tag{8}$$

which shows that any feasible solution to (6–7) indeed corresponds to a probability distribution. Therefore, we can define the partially identified set of interest as the set of solutions π satisfying

We can then derive analogous expressions to (3-4) in terms of upper contour sets:

$$p(U_{j,K})=\sum_{S\subseteq K}(-1)^{|S|}\Pr(j|(\mathcal{J}\setminus K)\cup S,p) \qquad \forall K\subseteq\mathcal{J},\forall j\notin K.$$

$$\Pr(j|B,p)=\sum_{\substack{K\supseteq\mathcal{J}\setminus B:\\ K\not\ni j}}p(U_{j,K}) \qquad \forall j\in\mathcal{J},\forall B\ni j.$$

(6–7). Define:

$$\mathcal{C} := \{C_{j,K} : K \subseteq \mathcal{J}, j \in \mathcal{J} \setminus K\} \cup \{\mathcal{R}\}$$
$$\mathcal{L} := \{L_{j,K} : K \subseteq \mathcal{J}, j \in \mathcal{J} \setminus K\} \cup \{\mathcal{R}\}.$$

When $\mathcal{S} = \mathcal{C}$, then (6-7) represents the partially identified set of possible probability distributions over strict preference orderings. This is because in this case the equations in (6) combined with (2) become:

$$\sum_{\succ \in C_{j,B\setminus j}} \pi_{\succ} = \Pr(j|B,p) \qquad\qquad \forall B \subseteq \mathcal{J}, \forall j \in B$$

This says that the sum of probabilities of preference orders that would select j from B is $\Pr(j|B,p)$, which is precisely what we observe in the data. So the partially identified set is convex and compact, and moreover it is a polytope (a bounded intersection of a finite number of half-spaces).

Similarly, if $\mathcal{S} = \mathcal{L}$, then it follows from Theorem 3 that (6-7) again describes the partially identified set, this time expressed in terms of lower contour sets instead of in terms of choice probabilities. One could also define the partially identified set in terms of upper contour sets.

To find extremal probabilities consistent with our observations, we consider two linear programs:

Maximum Probability (MaxP) Maximize $\sum_{\succ \in X} \pi_{\succ}$ subject to (6-7).

Minimum Probability (MinP) Minimize $\sum_{\succ \in X} \pi_{\succ}$ subject to (6-7).

In the above optimization problems, X can be any subset of $\mathcal{R}$. Thus, these are really optimization schema depending on the choice of $\mathcal{S}$ and X, encoding a variety of partial identification problems that find tight bounds on the probabilities of various sets of orderings. To give some examples, if X is the singleton $\{\succ'\}$ for some $\succ' \in \mathcal{R}$ and $\mathcal{S}$ is equal to either $\mathcal{C}$ or $\mathcal{L}$, then MaxP and MinP give, respectively, the maximum and minimum probabilities that can be assigned to the event that the consumer has preference ordering $\succ'$ consistent with the data. Alternatively, consider a subset K of $\mathcal{J}$, and let $\succ'$ be a preference ordering over the alternatives in K. Let X be the set of preference orderings on $\mathcal{J}$ that rank the alternatives in K as $\succ'$ does. Then with $\mathcal{S} = \mathcal{C}$ or $\mathcal{S} = \mathcal{L}$, MaxP and MinP give the maximum and minimum probabilities consistent with the data that the ordering of alternatives on the restricted subset K is given by $\succ'$. One could think of many other properties of preference whose extremal probabilities one might want to know and so could discern by this method.

There are also other candidate families for $\mathcal{S}$ that generate interesting bounds. For example, one could choose $\mathcal{S}$ to be a subset of $\mathcal{C}$ satisfying (5) and:

$$\forall B \subseteq \mathcal{J}, [\exists i \in B, C_{i,B\setminus i} \in \mathcal{S} \Rightarrow \forall j \in B, C_{j,B\setminus j} \in \mathcal{S}].$$

This encodes the statement that if we observe the choice probability corresponding to any item in B, we observe the choice probabilities corresponding to all items in B. This specification of $\mathcal{S}$ would be reasonable when we have not observed the distribution of choices from all budget sets but only from some subset of budget sets. Manski (2007) considered the special case of MaxP and MinP in which $\mathcal{S}$ satisfied these specifications and $X \in \mathcal{C} \setminus \mathcal{S}$.[4] In this case, X encodes the event that some element j would be chosen counterfactually from a budget set that we do not observe. This probability is constrained by the choice probabilities that we did observe. Alternatively, for any subset K of the budget set B that one does not observe, one could let X be the set of preference orders such that the consumer's selection from B would lie in K.

In all of the cases described above, finding the extremal probability consistent with the data amounts to a linear program, and hence any linear programming algorithm can be used to attain these probabilities. If $\mathcal{S}$ is equal to $\mathcal{C}$ or $\mathcal{L}$, this program may contain very many variables and constraints because the number of choice sets grows exponentially in the number of alternatives and the number of preference orders grow more than exponentially. When there is a small number of alternatives or when we have only observed choices from a limited number of choice sets, this program may be more manageable.

3.3 Degree of Underidentification of the Distribution of Preferences

It follows from Theorem 3 that the probability distribution p over orderings is identified when $\mathcal{J}$ contains three or fewer elements. In particular, suppose that $\mathcal{J} = \{1, 2, 3\}$. Consider the ordering $\succ$ such that

$$1 \succ 2 \succ 3. \tag{9}$$

Then observe that

$$p(\succ) = p(L_{2,\{3\}}) = \Pr(2|\{2,3\}, p) - \Pr(2|\{1,2,3\}, p).$$

The first equality follows from the fact that $\succ$ (given by (9)) is the unique ordering relation for which the strict lower contour set of 2 is the singleton $\{3\}$. The second equality–establishing identification–is a special case of equation (3) in Theorem 3. Similar equations hold for ordering relations other than (9). This argument fails when there are four or more choices.

Theorem 5. *When there are at least four alternatives, the mass function p is not identified from data on choice probabilities.*

Proof. Choose any probability distribution p^* with

$$p^*(\succ) > 0 \qquad \forall \succ \in \mathcal{R}$$

[4] The program considered by Manski is, however, more general in the sense that he allows for the possibility that agents make choices in an "irrational" way, in the sense that they choose in a way inconsistent with any preference ordering.

and consider the system of equations and inequalities:

$$\sum_{\succ \in L_{j,K}} \pi_{\succ} = p^*(L_{j,K}) \qquad \forall K \subseteq \mathcal{J}, \forall j \notin K \tag{10}$$

$$\sum_{\succ \in \mathcal{R}} \pi_{\succ} = 1 \tag{11}$$

$$\pi_{\succ} \geq 0 \qquad \forall \succ \in \mathcal{R}. \tag{12}$$

This system is just (6-7) when $\mathcal{S} = \mathcal{L}$. As mentioned above, part 2 of Theorem 3 implies that any feasible solution $\pi = (\pi_{\succ} : \succ \in \mathcal{R})$ to this system corresponds to a probability distribution over $\mathcal{R}$ that cannot be distinguished from p^*. By construction, (10-12) has at least one feasible solution π^*, namely the true distribution of preferences satisfying $\pi^*_{\succ} = p^*(\succ)$, $\forall \succ \in \mathcal{R}$. The system (10-11) contains $J2^{J-1} + 1$ equations and $J!$ unknowns. $J! > J2^{J-1} + 1$ as soon as $J \geq 5$, implying that (10-11) has infinitely many solutions. By construction, π^* satisfies all the conditions (12) with *strict* inequalities. Because one can find solutions π to (10-11) distinct from π^* but arbitrarily close to π^*, it is possible to find a solution π to (10-11) that is distinct from π^* and that also satisfies (12).

To extend this argument to the case $J = 4$, we argue that if $J \geq 3$, then at most $J(2^{J-1} - 3) + 3$ equations in (10-11) are linearly independent. This is proven as a lemma in the appendix. Observe finally that $J! > J(2^{J-1} - 3) + 3$ when $J = 4$. □

A typical demand estimation problem in the empirical literature considers only choice probabilities for consumers facing a single budget set B. The structural parameters of the model, p, will not be identified from data using a budget set B. The underidentification problem is rather severe. We can view equation (1) as a system of linear equations with $J!$ unknowns. The number of moments is easy to calculate. If a budget set B has n elements, then there are n moments corresponding to the choice probabilities $P(j \mid B, p)$, $j \in B$. If $n > 2$, $n \ll J!$. Not surprisingly, identification is hopeless without variation in B.

Now consider an econometrician who has access to data on choice probabilities for consumers with the same measure p facing different choice sets B. This could represent different geographic markets or the same market at different points in time. It could also represent experimental data. A necessary condition for the model to be identified is that the number of unknowns be no greater than the number of equations. The number of unknowns is still the probabilities of the $J!$ orderings. The number of moments for each budget size n is equal to the $\binom{J}{n}$ budgets B of size n times the number of products n. Observing all possible budgets of all possible sizes gives a total number of moments of

$$\sum_{n=2}^{J} n \binom{J}{n} = \frac{\left(2^J - 2\right) J!}{2\,(J-1)!} = \left(2^{J-1} - 1\right) J.$$

Table 1 lists the number of unknowns and market share moments and demonstrates that only models

Table 1: Number of Orderings and Moments

Choices	Orderings, $J!$	Moments, $\sum_{n=2}^{J} n\binom{J}{n}$
2	2	2
3	6	9
4	24	28
5	120	75
6	720	186
7	5040	441
8	40,320	1016
9	362,880	2295

with four or less choices can be identified using aggregate market shares, with all possible budget sets represented. In fact, as we show above, when $J = 4$ the model is also not identified because not all moments mentioned above are independent of one another. Lemma 12 in the appendix shows that when $J \geq 3$, at most $J(2^{J-1} - 3) + 3$ moments are independent. So the model is identified only when $J \leq 3$.

The degree of underidentification becomes quite severe as the number of choices increases. When there are 7 choices, there are 5040 probabilities on unknown orderings, but only 441 choice probabilities to match. This implies that the $J!$ structural primitives of our model are massively underidentified, even with hard to find aggregate data on the same mass of consumers facing different choice sets.

In any study that estimates the distribution of utility functions using data on choice probabilities for $J \geq 4$, identification will only arise from functional form assumptions and modeling observed product characteristics. This non-identification result and the numbers in Table 1 show that some restrictions are essential for the point identification of p in discrete choice analysis.

3.4 Identification of Preferences over Choice Sets

This section considers identification of the induced preferences over choice *sets*. In particular, we may be interested in the question of whether the substitution of some products for others will make consumers better or worse off. It would be desirable to rank choice sets to characterize the degree to which we can identify such rankings based on choice data. In this section, we show that when there are three choice sets under consideration, but possibly many alternatives within each of these possibly overlapping sets, then such a ranking is attainable from the data.

First consider the simpler case where there are only two budget sets A and B under consideration. Then we show that data on choice probabilities identifies the proportion of consumers who are better off under budget set A than under budget set B. This is the proportion of consumers who will vote for a particular budget set if given the choice between A and B. This in turn identifies the proportion of consumers who benefit from the addition, removal, or interchange of goods from a choice set. Thus, this result is useful for welfare analysis.

Any preference order $\succ$ over alternatives in $\mathcal{J}$ induces a preference order *sets* of alternatives. In particular for two budget sets A and B, we write $A \succ B$ if the $\succ$-best alternative in A is $\succ$-better than the $\succ$-best alternative in B. Even though we assume that $\succ$ is a *strict* preference order so that $\succ$ does not allow for any indifference between alternatives, $\succ$ may induce indifference between budget sets, written $A \sim B$. This is because the $\succ$-best alternative in A may be the same as the $\succ$-best alternative in B.

Let $\mathcal{B}$ be a family of subsets of $\mathcal{J}$. By the **distribution of preferences over** $\mathcal{B}$ we mean a probability distribution over the preference orderings over the budget sets within $\mathcal{B}$. As explained in the previous paragraph, any preference ordering over the alternatives in $\mathcal{J}$ induces a preference ordering over $\mathcal{B}$. So any probability distribution over preference orderings over $\mathcal{J}$ induces a probability distribution over preferences orderings over $\mathcal{B}$.

We use the notation $p(A \succ B)$ for the probability mass of consumers who prefer choice set A to choice set B, and $p(A \sim B)$ for the probability mass of consumers who are indifferent between A and B. We use similar notation for collections of three budget sets so that, for example, $p(A \succ B \sim C)$ is the probability mass of consumers who are indifferent between B and C, but prefer A to both B and C. Also, for any sets $A, B \subseteq \mathcal{J}$ with $A \subseteq B$, define

$$\Pr(A|B,p) = \sum_{j \in A} \Pr(j|B,p). \tag{13}$$

So $\Pr(A|B,p)$ is the probability that the choice alternative chosen from B will also lie in A. (13) implies that $\Pr(A|B,p)$ is identified from choice data. We now provide a positive result about the identification of the distribution of preferences over choice sets. Moreover, our result provides simple expressions for computing the desired probabilities.

Theorem 6. *If $|\mathcal{B}| \leq 3$, then the distribution of preferences over $\mathcal{B}$ is identified. More specifically:*

1. *If $\mathcal{B} = \{A, B\}$ (with $A \neq B$), then*

$$p(A \succ B) = \Pr(A \setminus B \mid A \cup B, p) \tag{14}$$

$$p(A \sim B) = \Pr(A \cap B \mid A \cup B, p). \tag{15}$$

2. *If $\mathcal{B} = \{A, B, C\}$ where A, B and C are distinct budget sets, then:*

$$p(A \succ B \succ C) = \Pr(B \setminus C \mid B \cup C, p) - \Pr(B \setminus C \mid A \cup B \cup C, p) \tag{16}$$

$$p(A \succ B \sim C) = \Pr(B \cap C \mid B \cup C, p) - \Pr(B \cap C \mid A \cup B \cup C, p) \tag{17}$$

$$p(A \sim B \succ C) = \Pr((A \cap B) \setminus C \mid A \cup B \cup C, p) \tag{18}$$

$$p(A \sim B \sim C) = \Pr(A \cap B \cap C \mid A \cup B \cup C, p). \tag{19}$$

Proof. (14) follows from the fact that a consumer prefers A to B if and only if when presented with $A \cup B$, she chooses an alternative in $A \setminus B$.

(15) follows from the fact that a consumer is indifferent between A and B if and only if when presented with $A \cup B$, she chooses an alternative in $A \cap B$.

Next we argue for (16). To do so, we argue that a consumer prefers A to B to C if and only if both (a) when presented with $B \cup C$ she chooses an alternative from $B \setminus C$ and (b) when presented with $A \cup B \cup C$ she no longer chooses an element of $B \setminus C$. Clearly if a consumer prefers A to B to C, she must satisfy both (a) and (b). Going in the other direction, (a) implies that the consumer prefers B to C. (b) then implies that when we add the alternatives in A to the set $B \cup C$, the consumer's choice changes, which implies that her favorite alternative is now in $A \setminus (B \cup C)$. So the consumer must prefer A to B to C.

Note finally that if a consumer satisfies the negation of (b), she must also satisfy (a), which explains the subtraction in (16).

(17) follows from the fact that a consumer prefers A to B and is indifferent between B and C if and only if both the consumer chooses an alternative from $B \cap C$ if presented with $B \cup C$ and no longer chooses an alternative from $B \cap C$ from $A \cup B \cup C$. The proof is similar to that of part (16).

(18) follows from the fact that a consumer is indifferent between A and B and prefers both A and B to C if and only if she chooses an element of $A \cap B$ from $A \cup B \cup C$.

(19) follows from the fact that a consumer is indifferent among A, B, and C if and only if she chooses an element of $A \cap B \cap C$ from $A \cup B \cup C$. $\square$

Unfortunately, identification of preferences over set of choices is not exhaustive when a researcher wants to identify preferences over four or more sets. In particular, Theorem 5 implies that when

$J = 4$ and $\mathcal{B} = \{\{1\}, \{2\}, \{3\}, \{4\}\}$, then preferences over $\mathcal{B}$ are not identified. We conjecture that the distribution of preferences over this family $\mathcal{B}$ are not identified even when $J > 4$. That is, we conjecture that the additional information provided in observing choice from sets including elements not in the set $\{1, 2, 3, 4\}$ would not allow us to point identify the distribution of preferences over $\{1, 2, 3, 4\}$.

3.5 Monotonicity In Product Characteristics Does Not Help

One approach to the point identification of p is to impose some structure on the choice set and preferences. For instance, suppose that an alternative can be evaluated along H dimensions, where a higher value along either dimension is known *a priori* to be better than a lower value. This imposes a limited knowledge of vectors of product characteristics x_j, but not a functional form as in $u_{i,j} = x_j'\beta_i + \varepsilon_{i,j}$, where β_i and $\epsilon_{i,j}$ are random terms, (e.g.) in the random coefficients logit.

For each of the H dimensions, suppose that each alternative may take a value in some finite set of integers values $\mathcal{V} = \{1, \ldots, V\}$, where higher values represent a higher quality along that dimension. The set of possible alternatives is then:

$$\mathcal{V}^H = \underbrace{\mathcal{V} \times \cdots \times \mathcal{V}}_{H \text{ times}}.$$

In other words, an alternative is determined by its value on each of a set of H dimensions. $\mathcal{V}^H$ plays the role of $\mathcal{J}$ in the previous analysis, so that we may write $\mathcal{J} = \mathcal{V}^H$.

In what follows, we restrict attention to preference orderings that are **monotone**, meaning that for all $(z_1, \ldots, z_H), (y_1, \ldots, y_H) \in \mathcal{V}^H$,

$$(z_1, \ldots, z_H) < (y_1, \ldots, y_H) \Rightarrow (z_1, \ldots, z_H) \prec (y_1, \ldots, y_H),$$

where $(z_1, \ldots, z_H) < (y_1, \ldots, y_H)$ means $z_h \leq y_h$ for all $h = 1, \ldots, H$ and $z_l < y_l$ for some $l = 1, \ldots, H$. This form of vector ordering imposes that the agent positively values all attributes but does not impose restrictions on how the agent trades off one characteristic versus another. Note also that we now have more data (observed product characteristics), but the preference orderings $\prec$ are now defined on all combinations of product characteristics. So both the data and the unknown distribution p are higher-dimensional objects than before.

Below we will also discuss the case where different attributes take on different sets of values so that for each attribute h the possible set of values is $\mathcal{V}_h = \{1, \ldots, V_h\}$. Note that given that attributes take on finitely many values, it is without loss of generality that the set of attributes is a set of consecutive integers starting with 1. We also discuss the case where the set of possible alternatives $\mathcal{J}$ is not equal to $\mathcal{V}^H$, but rather that $\mathcal{J} \subseteq \mathcal{V}^H$.

Definition 7. An **antichain** is a subset A of $\mathcal{V}^H$ such that for all $z, y \in A$ with $z \neq y$, it is neither the case that $z < y$ nor that $y < z$.

In words, an antichain is a collection of pairwise unordered bundles. Let $\alpha(H, V)$ be the maximum size antichain, which is a function of the number H of product characteristics and the number V of possible values taken on by each characteristic. Imposing monotonicity is not sufficient for the point identification of the distribution of preference orderings.

Theorem 8. *If $\alpha(H, V) > 3$, then the distribution of preferences p is not identified if all preference orderings are monotone.*

Proof. Let T be an antichain of size $\alpha(H, V)$ and define

$$S = \{x \in \mathcal{V}^H : \exists y \in T, x < y\}$$
$$U = \{z \in \mathcal{V}^H : \exists y \in T, y < z\}.$$

We now argue that $\{S, T, U\}$ is a partition of $\mathcal{V}^H$. This means that (i) the sets S, T, and U are pairwise disjoint and that (ii) $S \cup T \cup U = \mathcal{V}^H$. To establish (i), first note that the definitions of S and U and the fact that T is an antichain imply that $S \cap T = \emptyset$ and $U \cap T = \emptyset$. Next, assume for contradiction that there exists $y \in S \cap U$. Then the definitions of S and U imply that there exists $x \in T$ such that $x < y$ and there exists $z \in T$ such that $y < z$. But then $x < z$, contradicting the assumption that T is an antichain. This establishes (i). Next, assume for contradiction that there exists $x \in \mathcal{V}^H \setminus (S \cup T \cup U)$. It follows that $T \cup \{x\}$ is an antichain, contradicting the maximality of T. This establishes (ii).

Let $\succ_s$ (resp., $\succ_u$) be a monotone preference ordering on S (resp., U). Let $\mathcal{M}$ be the set of monotone preference relations on $\mathcal{V}^H$, and define

$$\mathcal{L} := \left\{ \succ \in \mathcal{M} : \forall x, y \in \mathcal{V}^{\mathcal{H}}, \left\{ \begin{array}{lc} x \in S \text{ and } y \in T & \text{or} \\ x \in S \text{ and } y \in U & \text{or} \\ x \in T \text{ and } y \in U & \text{or} \\ x, y \in S \text{ and } y \succ_s x & \text{or} \\ x, y \in U \text{ and } y \succ_u x & \end{array} \right\} \Rightarrow y \succ x \right\}. \tag{20}$$

By construction, $\mathcal{L} \subseteq \mathcal{M}$. We can associate with each preference $\succ$ over T the preference $\succ'$ in $\mathcal{L}$ that satisfies

$$x \succ' y \Leftrightarrow x \succ y \qquad \forall x, y \in T. \tag{21}$$

Because any preference comparison involving $\succ'$ is determined by (21) if $x, y \in T$, and by (20) otherwise, $\succ$ completely determines a preference ordering $\prec'$ in $\mathcal{L}$.

Because T is an antichain, monotonicity imposes no constraints on preferences over T. So insofar as we are interested in preferences over T, we can identify T with the set $\mathcal{J}' := \{1, \ldots, \alpha(H, V)\}$ through an arbitrary bijection f mapping $\mathcal{J}'$ onto T, and associate any preference relation over T with a preference relation over $\mathcal{J}'$. This induces a bijection between probability measures over preferences on $\mathcal{J}'$ and probability measures over preferences in $\mathcal{L}$ via (21). To see this, for any probability measure p over preferences over $\mathcal{J}'$, there is a probability measure p' over $\mathcal{L}$ satisfying $p(\succ) = p'(\succ')$, where $\succ'$ is the order in $\mathcal{L}$ corresponding to the order $\succ$ on $\mathcal{J}'$.

We now explore the properties a probability measure p' on $\mathcal{L}$ must have. For any $B \subseteq \mathcal{V}^H$ with $B \neq \emptyset$, we can consider three possibilities: (i) $B \cap U \neq \emptyset$, (ii) $B \subseteq S$, and (iii) $B \cap U = \emptyset, B \not\subseteq S$. In case (i), any probability measure p on $\mathcal{L}$ will be such that the unique $\succ_u$-maximal element in B will be chosen with probability 1. This element must be in U because of the definitions of U and $\mathcal{L}$. In case (ii), any probability measure p on $\mathcal{L}$ will be such that the unique $\succ_s$-maximal element in B will be chosen with probability 1, by the definition of $\mathcal{L}$. In case (iii), for any probability measure p' on $\mathcal{L}$, $\Pr(j|B, p') = 0$ for all $j \in B \setminus T$, and $\Pr(j|B, p') = \Pr(j|B \cap T, p')$ for all $j \in B \cap T$, by the definitions of T and $\mathcal{L}$. It follows that for a probability measure p' on $\mathcal{L}$, in order to describe $\Pr(j|B, p')$ for all $B \subseteq \mathcal{V}^H$, it is sufficient to specify $\Pr(j|B, p')$ for all $B \subseteq T$.

Given this, for any probability measure p on preference relations over $\mathcal{J}'$ and corresponding probability measure p' over $\mathcal{L}$, and $B \subseteq \mathcal{J}'$ and $j \in B$,

$$\Pr(j|B, p) = \Pr(f(j)|\{f(i) : i \in B\}, p'),$$

where f is the bijection introduced previously. It now follows from Theorem 5 that if $\alpha(H, J) > 3$, the distribution p' in the monotone random ordering model is not identified, because otherwise the distribution p would be identified in the non-monotone model when the set of alternatives is of size greater than 3. □

Corollary 9. *The distribution p over monotone preference orderings is not identified when any of*

- $H > 3$ *and* $V \geq 2$*, or*
- $H = 3$ *and* $V > 2$*, or*
- $H = 2$ *and* $V > 3$.

Proof. This follows from the fact that all three conditions imply the existence of an antichain which more than three elements. These facts can be algebraically verified. □

In other words the corollary says that with only a small number of characteristics (between two and four) and a small number of possible values per characteristic (between two and four), the distribution of preferences is not identified, even when we know that preferences are monotone in characteristics.

The problem only becomes more severe when the number of characteristics or the number of possible values per characteristic grows.

Remark 10. Theorem 8, which applies to sets of the form $\mathcal{V}^H$ under the componentwise ordering, can be generalized to arbitrary (finite) partially ordered choice sets. For any set of alternatives $\mathcal{J}'$, and an arbitrary partial ordering[5] $\leq$ on the set of alternatives $\mathcal{J}$, we can define a preference order $\prec$ to be $\leq$-monotone if:[6]

$$x < y \Rightarrow x \prec y \qquad\qquad \forall x, y \in \mathcal{J}. \tag{22}$$

Define a $\leq$-monotone probability distribution p to be a p where only preference orderings that are $\leq$-monotone receive positive probability. Then, using an argument essentially identical to the proof of Theorem 8, we can show that whenever $\leq$ allows for an antichain of size greater than three, the $\leq$-monotone probability distribution p is not identified.

Some special cases are worth mentioning. When each characteristic h may take on a different number of values V_h so that the choice set is of the form $\prod_{h\in H} \mathcal{V}_h$ where $\mathcal{V}_h = \{1, \ldots, V_h\}$, then the choice set is a partially ordered set, and so identification fails whenever there is an antichain of size greater than three. Likewise, if some combination of characteristics are impossible or never present so that $\mathcal{J}$ is a proper subset of $\mathcal{V}^H$, then again $\mathcal{J}$ is a partially ordered set, and identification fails whenever there is an antichain of size greater than three.

4 Prior Solutions for Point Identification

This section explores two sets of additional sources of data than can be brought to bear to restore point identification of p. These results are known in the literature, but referencing them here complements our non-identification results with solutions for point identification.

4.1 Panel Data on Individual Choices

Data on choice probabilities for the same population facing budget set variation are not enough to identify the population weights on the $J!$ orderings. Following the literature on economic theory, one solution is to consider panel data on individual consumers facing choice set variation. We assume that the consumers preferences remain constant over time, but the consumer may face a different choice set B_t in each period t. Let $j_{i,t}$ be the choice of agent i in period t. Importantly, compared to the previous section, we assume the data contain a panel identifier i so that we can track the consumer's

[5] Recall that a partial ordering is a reflexive, antisymmetric, and transitive relation.

[6] $<$ is the irreflexive part of $\leq$:

$$x < y \Leftrightarrow (x \leq y \text{ and } x \neq y)$$

.

choices over time. Let $(j_{i,1},...,j_{i,T})$ denote the string of choices of agent i in periods $t=1,\ldots,T$. For simplicity, we assume all agents face the same choice sets B_t in period t. Our data is now the probability of each string, or $\Pr(j_1,...,j_T|B_1,...,B_T,p)$. There is no identification gain from observing the same budget set twice, so we exclude duplicate budget sets.

Reasoning as in the previous linear-in-probabilities model (1) it must be the case that

$$\Pr(j_1,...,j_T|B_1,...,B_T,p)=\sum_{\succ\in\mathcal{R}}p(\succ)\,1\{\text{for all } t,\ j_t\succ j_t' \text{ for all } j_t'\in B_t\}. \tag{23}$$

Notice how the number of moments has just dramatically increased. With aggregate market shares we have at most $\sum_{n=2}^{J}n\binom{J}{n}=\left(2^{J-1}-1\right)J$ moments. With individual panel data with exhaustive budget sets, we have an upper bound of $\prod_{n=2}^{J}n\binom{J}{n}$ strings of choices, as a string of choices is one action from each budget set.

In reality, there are exactly $J!$ possible strings that could be generated by rational choices, as each preference ordering corresponds to a different string of choices. So there will be a one-to-one correspondence between the $J!$ empirically-observable strings and the $J!$ unknown probabilities captured by $p(\succ)$. As is known from classical choice theory, each individual i's ordering can be recovered. Indeed, observing all budget sets with two items each is enough to identify a transitive preference relation (Richter 1971; Mas-Colell, Whinston, and Green 1995, Chapter 1).

There is another approach we may take to identify p using cross sectional data, which has some conceptual similarities to panel data. We can use surveys where individuals report their entire rankings. If the economist sees, for instance, the entire ranking instead of the demanded bundle, the distribution over preference orderings is trivially identified. We note that such surveys are common in conjoint analysis and applied marketing.

4.2 Quasi-Linear Utility and Price Variation

Section 3 showed that data on choice probabilities, even with choice set variation, is not sufficient to identify p. Individual panel data with choice set variation is one solution to this problem. We next turn to identification through restricting preferences and price variation.

In random utility models, such as those reviewed in the introduction and Section 2, it is common to assume that the contribution of the (non-price) product characteristics of j is additively separable from the contribution of a composite commodity to utility, representing income spent on goods outside the category under study. We show that if analogous restrictions are made on utility, then the distribution of consumer preferences is identified with price variation over time.

We model utility from a choice as additively separable into a base component $u_{i,j}$ and the value of the composite component c. In this case we redefine the commodity space to be $\mathcal{J}:=\{1,\ldots,J\}\times\mathbb{R}$, where the second component (in $\mathbb{R}$) represents the consumption of the composite commodity. $\mathcal{R}$ now represents a set of preference relations over this enlarged commodity space. Assume that $\mathcal{R}$ is restricted to those preference relations that can be represented by a utility function that is quasi-linear in c and that the marginal utility of consumption is strictly positive. That is, each $\succ_i \in \mathcal{R}$ can be represented by a utility function of the form $u_{i,j} + c$, where $u_{i,j}$ is a mapping from the set $\{1, ..., J\}$ to the reals.

Here the additive separability of the composite commodity is the same as the transferable utility assumption in the literature on firm behavior in markets, for example. We do not include a coefficient on c because of an innocuous scale normalization in the value of utilities, in order to rule out having two utility functions that give the same preference ordering. Likewise, we also impose a location normalization that $u_{i,1} = 0$ for good 1 for all consumers i. This often represents the option of not purchasing a good in a particular category; it otherwise is a normalization.

Each product j has a price q_j, which varies across choice situations. Also, each consumer has an income y_i. Substituting these terms into $u_{i,j} + c$ gives $u_{i,j} + y_i - q_j$. By the discrete choice decision rule, y_i will not affect choices. As the units of utility are arbitrary, we might as well express them in monetary units, so $u_{i,j}$ represents the willingness to pay for product j. The goal of the empirical work is to use price variation to identify the distribution $F_u(u_2, \ldots, u_J)$ of the $u_{i,j}$'s. There are still $J!$ orderings of the $u_{i,j}$'s, but the introduction of prices makes the distribution $F_u(u_2, \ldots, u_J)$ of now-cardinal willingness to pay the object of identification. Because of our cardinalization of utility, the same preference ordering $\succ$ over non-price characteristics can lead to different willingnesses to pay.

While specifying utility as quasi-linear in the composite commodity is restrictive, the assumptions is commonly made in the literature. As noted in Anderson, DePalma, and Thisse (1992), the quasi-linear assumption is useful as it allows the economist to recover consumers preferences without data on income. The assumption of quasilinear utility assumes that the marginal utility of the composite commodity is constant. In many commonly studied product categories, such as consumer packaged goods, this is probably a fairly reasonable approximation since the purchase price will be small compared to income.

Unlike many of the papers mentioned in the introduction, we do not decompose the willingness to pay $u_{i,j}$ into observed covariates and error terms, such as $u_{i,j} = \beta_{1,i}x_{1,j} + \ldots + \beta_{k,i}x_{k,j} + \varepsilon_{i,j}$. We do not require these combinations of functional forms and independence assumptions on the random coefficients for identification, as is shown below.

Based on the arguments similar to those in the previous section, it is easy to see that flexible price variation will identify the vector of $u_{i,j}$'s using panel data on individual choices for consumer i. We shall instead address estimating the distribution $F_u(u_2, \ldots, u_J)$ using only aggregated choice probabilities or market shares, which corresponds to cross-sectional data on individual choice.

Our approach of using price variation is equivalent to the special regressor approach to identification in the econometrics literature. See for example, Lewbel (2000), Matzkin (2007), Berry and Haile (2010) and Fox and Gandhi (2010) for a few of many examples of special regressor arguments in multinomial choice. As with these other papers, we need the vector of J prices q to have full support on $\mathbb{R}^J$ if the utility functions are unrestricted. We make that assumption. Observing negative prices is only an issue if negative utilities need to be identified.

Theorem 11. *The distribution $F_u(u_2, \ldots, u_J)$ of willingnesses to pay is identified if the vector of J prices q has full support on $\mathbb{R}^J$.*

Proof. We follow the simple argument of Berry and Haile (2010). The choice probability of good 1 at a vector of prices is

$$\Pr(1|q) = F_u(q_2 - q_1, \ldots, q_J - q_1),$$

as good 1 is purchased whenever

$$0 - q_1 > u_j - q_j \text{ or } u_j < q_j - q_1 \ \forall j = 2, \ldots, J.$$

Thus, choice probabilities for each q identify the distribution F_u at each point of its evaluation. □

If sufficient price variation does not exist in the data, then the distribution $F_u(u_2, \ldots, u_J)$ will be identified only on its arguments corresponding to prices in the data. This can be seen as a form of set identification.

5 Conclusion

We consider the identification of the distribution of heterogeneity in discrete choice models, using cross-sectional data on choice probabilities or market shares. Compared to the prior empirical literature, we do not model the choices as lying in a product space of observable product characteristics. We model individuals' preferences using a preference ordering, which strictly ranks a finite set of alternatives. The distribution of preference orderings is generally only partially identified from data on choice probabilities at all budget sets if the number of choices is greater than three. We numerically characterize the degree of underidentification. Then using a linear programming approach to partial identification, we show how to obtain bounds on probabilities of any ordering relation consistent with the data. This approach can also produce bounds on probabilities for various properties of preferences, including counterfactual choice probabilities when not all choice situations are observed.

While the distribution of preference orderings is generally only partially identified, an important feature of demand is point identified from choice probabilities, however. We can identify consumers' voting preferences: what proportion of consumers prefer a particular budget set B instead of a competing budget set B'.

The general point-identification result for random utility models implies that identification of these models must rely on a space of observable product characteristics and other modeling assumptions. We establish that imposing monotonicity restrictions on preferences does not restore point identification of the distribution of preferences, however. Using individual level panel data on choices, we show that the distribution of preferences is point identified. Also, a quasi-utility specification with enough price variation can deliver point identification of the distribution of willingnesses to pay.

A Appendix

A.1 Proof of Theorem 3

First we prove part 1. We start with the set-theoretic relation

$$L_{j,K} = C_{j,K} \setminus \bigcup_{i \in \mathcal{J} \setminus (K \cup j)} C_{j,K \cup i}, \qquad \forall K \subseteq \mathcal{J}, \forall j \notin K. \tag{24}$$

This follows because under $\succ$, K is the lower contour set of j if and only if K is contained in the lower contour set of j (i.e., $\succ \in C_{j,K}$), but for all $i \notin (K \cup j)$, $K \cup i$ is not contained in the lower contour set of j (i.e., $\succ \notin C_{j,K \cup i}$). Using an inclusion-exclusion formula, we have:

$$\begin{aligned} p(\bigcup_{i \in \mathcal{J} \setminus (K \cup j)} C_{j,K \cup i}) &= \sum_{s=1}^{n(K)} (-1)^{s-1} \sum_{\substack{S \subseteq \mathcal{J} \setminus (K \cup j): \\ |S| = s}} p(\bigcap_{i \in S} C_{j,K \cup i}) && (25) \\ &= \sum_{s=1}^{n(K)} (-1)^{s-1} \sum_{\substack{S \subseteq \mathcal{J} \setminus (K \cup j): \\ |S| = s}} p(C_{j,K \cup S}) && (26) \end{aligned}$$

where:

$$n(K) := |\mathcal{J} \setminus (K \cup j)| = |\mathcal{J} \setminus K| - 1$$

(26) follows from the set-theoretic equality:

$$\bigcap_{i \in S} C_{j,K \cup i} = C_{j,K \cup S}.$$

This equality holds because $\bigcup\{K \cup i : i \in S\} = K \cup S$, so that $K \cup i$ is contained in the lower contour set of j for all $i \in S$ if and only if $K \cup S$ is contained in the lower contour set of j.

Next observe that for all $i \in \mathcal{J} \setminus (K \cup j)$, $C_{j,K} \supseteq C_{j,K \cup i}$ because if $K \cup i$ is contained in the lower contour set of j, then so is K. It follows that

$$C_{j,K} \supseteq \bigcup_{i \in \mathcal{J} \setminus (K \cup j)} C_{j,K \cup i}. \tag{27}$$

So,

$$
\begin{aligned}
p(L_{j,K}) &= p(C_{j,K} \setminus \bigcup_{i \in \mathcal{J} \setminus (K \cup j)} C_{j,K \cup i}) \\
&= p(C_{j,K}) - p(\bigcup_{i \in \mathcal{J} \setminus (K \cup j)} C_{j,K \cup i}) \\
&= p(C_{j,K}) - \sum_{s=1}^{n(K)} (-1)^{s-1} \sum_{\substack{S \subseteq \mathcal{J} \setminus (K \cup j) : \\ |S| = s}} p(C_{j,K \cup S}) \qquad (28) \\
&= \sum_{s=0}^{n(K)} (-1)^{s} \sum_{\substack{S \subseteq \mathcal{J} \setminus (K \cup j) : \\ |S| = s}} p(C_{j,K \cup S}),
\end{aligned}
$$

where the first equality follows from (24), the second equality follows from (27), and the third equality follows from (25-26). Note finally that, using (2), (28) is equivalent to (3), establishing part 1 of the theorem.

Part 2 of the theorem follows from (2) and the fact that the family of sets

$$\mathcal{P} := \{L_{j,K} : K \supseteq B \setminus j, K \not\ni j\}$$

is a partition of the set $C_{j,K}$. The latter fact holds because if the lower contour set of j contains $B \setminus j$, then the lower contour set of j is some superset of $B \setminus j$ not containing j, and $\mathcal{P}$ contains all of the (mutually exclusive) possibilities.

A.2 Proof of Lemma Supporting Theorem 5

Lemma 12. *If $J \geq 3$, then at most $J(2^{J-1}-3)+3$ of the equations in (10-11) are linearly independent.*

Proof. Define $H^j := \{1, 2, \ldots, j-1\}$, and consider the sets:

$$
\begin{aligned}
\mathcal{B}_1 &:= \{(j, H^j) : j = 2, \ldots, J-1\} \\
\mathcal{B}_2 &:= \{(j, \mathcal{J} \setminus j) : j = 1, \ldots, J\} \\
\mathcal{B}_3 &:= \{(j, \emptyset) : B = \emptyset, j = 1, \ldots, J\}.
\end{aligned}
$$

Observe that the sets $\mathcal{B}_1, \mathcal{B}_2$, and $\mathcal{B}_3$ are pairwise disjoint (here we use the fact that $J \geq 3$). Finally define:

$$\mathcal{B}_0 := \{(j, B) : B \subseteq \mathcal{J}, j \notin B, (j, B) \notin \mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{B}_3\}$$

Consider the system of equations:

$$\sum_{\succ \in L_{j,B}} \pi_\succ = p^*(L_{j,B}) \qquad \forall (j,B) \in \mathcal{B}_0 \tag{29}$$

$$\sum_{\succ \in \mathcal{R}} \pi_\succ = 1. \tag{30}$$

It is straightforward to confirm that (29–30) contains $J(2^{J-1} - 3) + 3$ equations. To complete the proof, we now argue that any solution $\pi = (\pi_\succ : \succ \in \mathcal{R})$ to (29–30) is a solution to (10–11). So let π be a solution to (29–30).

1. First choose H^j for $j = 2, \ldots, J-1$. We have:

$$\begin{aligned} p^*(L_{j,H^j}) &= 1 - \sum_{B \subseteq \mathcal{J} \setminus j : |B| = j-1, B \neq H^j} p^*(L_{j,B}) - \sum_{\{i \in \mathcal{J} \setminus j\}} \sum_{\{A \subseteq \mathcal{J} \setminus i : |A| = j-1\}} p^*(L_{i,A}) \\ &= 1 - \sum_{\{B \subseteq \mathcal{J} \setminus j : |B| = j-1, B \neq H^j\}} \sum_{\{\succ \in L_{j,B}\}} \pi_\succ - \sum_{\{i \in \mathcal{J} \setminus j\}} \sum_{\{A \subseteq \mathcal{J} \setminus i : |A| = j-1\}} \sum_{\{\succ \in L_{j,A}\}} \pi_\succ \\ &= \sum_{\succ \in L_{j,H^j}} \pi_\succ . \end{aligned} \tag{31}$$

   The first and last equalities follows from the fact that

$$\begin{aligned} \mathcal{P} :=& \{L_{j,H^j}\} \cup \{L_{j,B} : B \subseteq \mathcal{J} \setminus j, |B| = j-1, B \neq H^j\} \\ & \cup \{L_{i,A} : i \in \mathcal{J} \setminus j, A \subseteq \mathcal{J} \setminus i, |A| = j-1\} \end{aligned}$$

   is a partition of $\mathcal{R}$. $\mathcal{P}$ is a partition because it specifies every possible collection of the bottom $j-1$ ranked items, as well the item ranked jth from the bottom. The last equality also uses (30). The second equality in (31) follows from (29) and the fact that all cells $L_{i,B}$ of the partition $\mathcal{P}$ except L_{j,H^j} are such that $(i,B) \in \mathcal{B}_0$.

2. Next choose $j = 1, \ldots, J$. We have

$$p^*(L_{j,\mathcal{J}\setminus j}) = \sum_{i \in \mathcal{J} \setminus j} p^*(L_{i,\mathcal{J}\setminus\{i,j\}}) = \sum_{\{i \in \mathcal{J} \setminus j\}} \sum_{\{\succ \in L_{i,\mathcal{J}\setminus\{i,j\}}\}} \pi_\succ = \sum_{\succ \in L_{j,\mathcal{J}\setminus j}} \pi_\succ . \tag{32}$$

   The first and last equalities follow from the fact that

$$\mathcal{P}' := \{L_{i,\mathcal{J}\setminus\{i,j\}} : i \in \mathcal{J} \setminus j\}$$

   is a partition of $L_{j,\mathcal{J}\setminus j}$ because $\mathcal{P}'$ specifies all possible 2nd ranked items when the 1st ranked item is j. Notice that for all $L_{i,B} \in \mathcal{P}'$ if $i = J-1$ and $j = J$, then $B = H^j$ and $(i,B) \in \mathcal{B}_1$, but otherwise $(i, B) \in \mathcal{B}_0$. So the second equality in (32) follows from (30) and (31).

3. Finally consider $j = 1, \ldots, J$. We have

$$p^*(L_{j,\emptyset}) = 1 - \sum_{B \subseteq \mathcal{J} \setminus j : B \neq \emptyset} p^*(L_{j,B}) = 1 - \sum_{\{B \subseteq \mathcal{J} \setminus j : B \neq \emptyset\}} \sum_{\{\succ \in L_{j,B}\}} \pi_{\succ} = \sum_{\succ \in L_{j,\emptyset}} \pi_{\succ}. \tag{33}$$

   The first and last equalities follow from (30) and the fact that

$$\mathcal{P}'' := \{(j, B) : B \subseteq \mathcal{J} \setminus j\}$$

   is a partition of $\mathcal{R}$ because $\mathcal{P}''$ specifies every possible strict lower contour set of j. The second equality in (33) follows from (29), (31) and (32).

Arguments 1–3 imply that any feasible solution of (29–30) is a feasible solution of (10–11), completing the proof. □

References

[1] Anderson, S.P., A. de Palma, and J.F. Thisse (1992), *Discrete Choice Theory of Product Differentiation*, The MIT press.

[2] Barbera, S. and P. Pattanaik (1986), "Falmagne and the Rationalizability of Stochastic Choices in Terms of Random Orderings ", *Econometrica*, 54, 707–715.

[3] Bajari, P., J. Fox, K. Kim, and S. Ryan (2010), "The Random Coefficient Logit Model is Identified ", University of Minnesota working paper.

[4] Berry, S., J. Levinsohn, and A. Pakes (1995), "Automobile Price in Market Equilibrium", *Econometrica* (63), July 1995.

[5] Berry, S. and P. Haile (2010), "Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers", Yale University working paper.

[6] Block, H. D. and J. Marschak (1960), "Random Orderings and Stochastic Theories of Responses". in I. Olkin, S. Ghurye, W. Hoeffding, W. Madow, and H. Man (eds) *Contributions to Probability and Statistics*, Stanford University Press.

[7] Chiappori, P.A. and I. Komunjer (2009), "On the Nonparametric Identification of Multiple Choice Models", Columbia University working paper.

[8] Falmagne, J.C. (1978), "A Representation Theorem for Finite Random Scale Systems", *Journal of Mathematical Psychology*, 18, 52–72.

[9] Fishburn, P. (1992) "'Induced binary probabilities and the linear ordering polytope: a status report", *Mathematical Social Sciences*, 23(1), 67–80.

[10] Fishburn, P. (1998) "Stochastic Utility ", in S. Barbara, P. Hammond, C. Seidl (eds) *Handbook of Utility Theory*, Kluwer, 273-320.

[11] Goettler, R. (1999), "Advertising Rates, Audience Composition, and Competition in the Network Television Industry ", Carnegie Mellon University working paper.

[12] Fox, J. and A. Gandhi (2010), "Nonparametric Identification and Estimation of Random Coefficients in Nonlinear Economic Models", University of Michigan working paper.

[13] Hitsch, G.J. (2006), "An empirical model of optimal dynamic product launch and exit under demand uncertainty", *Marketing Science*, 25(1), 25–50.

[14] Ichimura, H. and Thompson, T.S. (1998), "Maximum likelihood estimation of a binary choice model with random coefficients of unknown distribution", *Journal of Econometrics*, 86(2), 269–295.

[15] Lewbel, A. (2000), "Semiparametric Qualitative Response Model Estimation with Unknown Heteroskedasticity or Instrumental Variables", *Journal of Econometrics*, 97(1), 145-177.

[16] Mas-Colell, A., M.D. Whinston, and J. Green (1995), *Microeconomic Theory*, Oxford University Press.

[17] Manski, C.F. (1977), "The Structure of Random Utility Models", *Theory and Decisions*, 8, 229–254.

[18] Manski, C.F. (1995), *Identification Problems in the Social Sciences, Cambridge*, MA: Harvard University Press.

[19] Manski, C.F. (2003), *Partial Identification of Probability Distributions*, New York: Springer-Verlag.

[20] Manski, C.F. (2007), "Partial Identification of Counterfactual Choice Probabilities", *International Economic Review,* 48(4), 1393–1410.

[21] Matzkin, R.L. (1993) "Nonparametric Identification and Estimation of Polychotomous Choice Models", *Journal of Econometrics*, 58, 137–168.

[22] Matzkin, R.L. (2007) "Heterogeneous Choice", *Advances in Economics and Econometrics, Theory and Applications*, Ninth World Congress of the Econometric Society.

[23] McFadden, D. (1972), "Conditional Logit Analysis of Qualitative Choice Behavior", *Frontiers in Econometrics*, 105–142, Academic Press.

[24] McFadden, D. (2005), "Revealed Stochastic Preference: A Synthesis", *Economic Theory*, 26(2), 245–264.

[25] McFadden, D. and K. Richter (1991) "Stochastic Rationality and Revealed Stochastic Preference", in J. Chipman, D. McFadden, K. Richter (eds) *Preferences, Uncertainty, and Rationality*, Westview Press, 161–186.

[26] Nevo, A (2001), "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica*, 69(2), 307–342.

[27] Petrin, A. (2002), "Quantifying the Benefits of New Products: The Case of the Minivan," *Journal of Political Economy*, 110(4), 705–729.

[28] Richter, M. (1971), "Rational Choice", in J. Chipman, L. Hurwicz, and H. Sonnenschein (eds) *Preferences, Utility, and Demand*, Harcourt Brace Jovanovich, Chapter 2.