UNIVERSITY OF CAMBRIDGE
Faculty of Economics

Cambridge Working Papers in Economics

Cambridge Working Papers in Economics: 1884

Income Effects and Rationalizability in Multinomial Choice Models

Debopam Bhattacharya

1 August 2018

In multinomial choice settings, Daly-Zachary (1978) and Armstrong-Vickers (2015) provided closedform conditions, under which choice probability functions can be rationalized via random utility models. A key condition is Slutsky symmetry. We first show that in the multinomial context, Daly-Zachary's Slutsky symmetry is equivalent to absence of income-effects. Next, for general multinomial choice that allows for income-effects, we provide global shape restrictions on choice probability functions, which are shown to be sufficient for rationalizability. Finally, we outline nonparametric identification of preference distributions using these results. The theory of linear partial differential equations plays a key role in our analysis.

Income Effects and Rationalizability in Multinomial Choice Models

Debopam Bhattacharya

University of Cambridge

August, 2018

Abstract

In multinomial choice settings, Daly-Zachary (1978) and Armstrong-Vickers (2015) provided closed-form conditions, under which choice probability functions can be rationalized via random utility models. A key condition is Slutsky symmetry. We first show that in the multinomial context, Daly-Zachary's Slutsky symmetry is equivalent to absence of income-effects. Next, for general multinomial choice that allows for income-effects, we provide global shape restrictions on choice probability functions, which are shown to be sufficient for rationalizability. Finally, we outline nonparametric identification of preference distributions using these results. The theory of linear partial differential equations plays a key role in our analysis.

Keywords: Multinomial Choice, Unobserved Heterogeneity, random Utility, Rationalizability/Integrability, Slutsky-Symmetry, Income Effects, Partial Differential Equations, Nonparametric Identification.

JEL Codes: C14, C25, D11.

1 Introduction

The random utility model of multinomial choice (McFadden, 1973) has gained immense popularity among applied economists. However, there has been limited research on the micro-theoretic underpinning of such models, and in particular, on the question of which choice probability functions are logically consistent with a random utility model.[1] Daly and Zachary 1978 provided a set of closed-form, global conditions under which choice-probability *functions* can be justified as having arisen from preference maximization by a heterogeneous population. These conditions were re-stated in Anderson et al, 1992, Theorem 3.1, and independently derived in Armstrong and Vickers, 2015, who improved upon the Daly-Zachary results by

[1] There has been comparatively more work on rationalizability in empirical demand models with *continuous* goods, c.f. Lewbel (2001).

including an outside option in the choice set. In all of these results, a key condition for rationalizability is Slutsky symmetry, analogous to the classic textbook case for demand systems with continuous goods.

In this paper, we first show that in the multinomial setting, Daly-Zachary's Slutsky symmetry is equivalent to the absence of income effects, i.e. that conditional choice probabilities do not depend on the decision-makers' income. The "necessity" part is easy to show. Showing "sufficiency", i.e. that Slutsky symmetry implies absence of income effects is non-trivial, and represents the first result of the paper. Next, in multinomial settings that allow for income effects, we provide a set of alternative shape restrictions on conditional choice probability functions, including a counterpart of Slutsky symmetry, which together are shown to be sufficient for rationalizability. The proof of this result is constructive, and the rationalizing utility functions are obtained by inverting solutions of certain linear partial differential equations (PDEs). The way in which PDEs arise here is unrelated to Roy's Identity (c.f. Mas-Colell et al, 1995, Proposition 3.G.4); in particular, the partial derivatives appearing in the PDE are of the average *demand* function, not the indirect utility function.

Finally, we show that the rationalizability results can also be used to nonparametrically identify the underlying preference distributions from empirical choice-probabilities. A key restriction delivering this identification result – viz. invertibility of sub-utilities in the numeraire due to non-satiation – is based on *economic* theory, as opposed to statistical assumptions. Furthermore, achieving nonparametric identification by solving PDEs appears to be novel in the discrete choice literature.

The plan for the rest of the paper is as follows. Section 2 discusses Daly-Zachary's Slutsky symmetry condition, and its connection with lack of income effects. Section 3 discusses rationalizability for multinomial choice in presence of income effects, and presents Theorem 1, the key result of this paper. Section 4 discusses some further points, including the implication of the rationalizability result for nonparametric identification of preference distributions. A short appendix at the end presents two mathematical results on partial and ordinary differential equations that are intensively used in the paper.

2 The Daly-Zachary Result

Consider a setting of multinomial choice, where the discrete alternatives are indexed by $j = 0, 1, ..., J$, individual income is y, price of alternative j is p_j; if alternative 0 refers to the outside option, i.e. not buying any of the alternatives, then $p_0 \equiv 0$. Let the utility from consuming the jth alternative and a quantity z of the numeraire be given by $U(j,z)$. The consumer's problem is $\max_{j\in\{0,1,...,J\},z}[U(j,z)+\varepsilon_j]$, subject to the budget constraint $z \leq y - p_j$, where y is the consumer's income, p_j is the price of alternative j faced by the consumer, and ε_j is unobserved heterogeneity in the consumer's preferences. If $U(j,\cdot)$ is strictly increasing (i.e. non-satiation in the numeraire), then we can replace $z = y - p_j$, and rewrite the consumer problem as $\max_{j\in\{0,1,\ldots,J\}}[U(j,a_j)+\varepsilon_j]$, where $a_j \equiv y - p_j$. Denote the (structural) probability of choosing alternative $j \in \{0,\ldots,J\}$ at $\mathbf{a} \equiv (a_0,..,a_J)$ by $q_j(\mathbf{a})$. In words, if we randomly sample individuals from the population, and offer the vector $\mathbf{a}$ to each sampled individual, then a fraction $q_j(\mathbf{a})$ will choose alternative j, in expectation. It is easy to incorporate other attributes of the alternatives and characteristics of consumers in our analysis, and we outline how to that below, after Theorem 1. For now, we suppress other covariates for clarity of exposition.

Slutsky-Symmetry: In this set-up, Daly-Zachary's Slutsky symmetry conditions are that for any two alternatives $k,l \in \{0,1,\ldots,J\}$, $k \neq l$,

$$\frac{\partial}{\partial a_l}q_k(\mathbf{a}) = \frac{\partial}{\partial a_k}q_l(\mathbf{a}).\text{[2]} \tag{1}$$

We first show that the classic random utility model with no income effects implies (1). We then show the first result of our paper, viz. that Slutsky symmetry (1) implies absence of income effects.

Necessity: The canonical random utility model of multinomial choice assumes that utility from consuming the jth alternative at income y and price p_j is given by

$$U(j,a_j) = a_j, \tag{2}$$

where $a_j = y - p_j$ as above. Income effects are zero since demand depends on the a's via the differences $a_j - a_k = (y-p_j)-(y-p_k) = p_k - p_j$. Suppose $(\varepsilon_0,\varepsilon_1,\ldots\varepsilon_J)$ are continuously distributed with joint density $g(\cdot)$. Then, the choice probability for the 0th alternative is given by

$$\begin{aligned} & q_0(\mathbf{a}) \\ = \; & \Pr\left(\cap_{j\neq 0}\left\{a_0+\varepsilon_0 > a_j+\varepsilon_j\right\}\right) \\ = \; & \Pr\left(\cap_{j\neq 0}\left\{a_0-a_j > \varepsilon_j-\varepsilon_0\right\}\right) \\ = \; & \int_{-\infty}^{\infty}\int_{-\infty}^{a_0-a_1+\varepsilon_0}\ldots\int_{-\infty}^{a_0-a_J+\varepsilon_0} g(\varepsilon)\,d\varepsilon_J\ldots d\varepsilon_1 d\varepsilon_0. \end{aligned}$$

[2]Daly-Zachary defines choice probabilities as functions of price and income, $\tilde{q}_j(p_0,p_1,\ldots,p_J,y)$. This is equivalent to our notation of $q_j(a_0,a_1,\ldots a_J)$ with $a_0 = y$, $a_1 = y-p_1$,...,$a_J = y-p_J$, in that one can move back and forth between the two notations, since

$$\begin{aligned} q_j(a_0,a_1,\ldots,a_J) &= \tilde{q}_j(a_0-a_1,a_0-a_2,\ldots,a_0-a_J),\text{ and} \\ \tilde{q}_j(p_1,p_2,\ldots,p_J,y) &= q_j(y,y-p_1,y-p_2,\ldots,y-p_J). \end{aligned}$$

"Slutsky symmetry" in Daly-Zachary's notation is that $\partial\tilde{q}_k/\partial p_j = \partial\tilde{q}_j/\partial p_k$ for all $j \neq k$ (if alternative 0 is the ouside option, then the corresponding condition is $\partial\tilde{q}_0/\partial p_j = \partial\tilde{q}_j/\partial y$). which is identical to (1) in our notation.

Therefore, by the first fundamental theorem of calculus,

$$\begin{aligned}&\frac{\partial}{\partial a_1}q_0(\mathbf{a})\\ =\ & -\int_{-\infty}^{\infty}\int_{-\infty}^{\varepsilon_0+a_0-a_2}\ldots\int_{-\infty}^{\varepsilon_0+a_0-a_J} g(\varepsilon_0, a_0-a_1+\varepsilon_0, \varepsilon_2, \ldots\varepsilon_J)\, d\varepsilon_J\ldots d\varepsilon_2 d\varepsilon_0. \qquad (3)\end{aligned}$$

Similarly,

$$q_1(\mathbf{a}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\varepsilon_1+a_1-a_0}\ldots\int_{-\infty}^{\varepsilon_1+a_1-a_J} g(\varepsilon)\, d\varepsilon_J\ldots d\varepsilon_2 d\varepsilon_0 d\varepsilon_1,$$

implying

$$\begin{aligned}&\frac{\partial}{\partial a_0}q_1(\mathbf{a})\\ =\ & -\int_{-\infty}^{\infty}\int_{-\infty}^{a_1-a_2+\varepsilon_1}\ldots\int_{-\infty}^{a_1-a_J+\varepsilon_1} g((a_1-a_0+\varepsilon_1, \varepsilon_1, \varepsilon_2, \ldots\varepsilon_J|, \gamma)\, d\varepsilon_J\ldots d\varepsilon_2 d\varepsilon_1 \qquad (4)\\ =\ & -\int_{-\infty}^{\infty}\int_{-\infty}^{s_0+a_0-a_2}\ldots\int_{-\infty}^{s_0+a_0-a_J} g(s_0, s_0-a_1+a_0, \varepsilon_2, \ldots\varepsilon_J)\, d\varepsilon_J\ldots d\varepsilon_2 ds_0\\ =\ & \frac{\partial}{\partial a_1}q_0(\mathbf{a}), \text{ using (3)},\end{aligned}$$

where the second equality follows by substituting $s_0 = a_1 - a_0 + \varepsilon_1$ in (4).

The same argument can be repeated for any other pair of alternatives $l \neq k$, to obtain

$$\frac{\frac{\partial}{\partial a_k}q_l(\mathbf{a})}{\frac{\partial}{\partial a_l}q_k(\mathbf{a})} = 1, \qquad (5)$$

for all $\mathbf{a}$. This shows that in the canonical random utility model with no income effects, Daly-Zachary's Slutsky symmetry condition holds.

Sufficiency: We now show that Slutsky symmetry implies absence of income effects. To see this, first note that because $\sum_{k=0}^{J} q_k(\mathbf{a}) = 1$, differentiating both sides w.r.t. a_l gives

$$\frac{\partial}{\partial a_l}q_l(\mathbf{a}) + \sum_{k=0,k\neq l}^{J}\frac{\partial}{\partial a_l}q_k(\mathbf{a}) = 0. \qquad (6)$$

Substituting (1) in (6), we get:

$$\frac{\partial}{\partial a_l}q_l(\mathbf{a}) + \sum_{k=0,k\neq l}^{J}\frac{\partial}{\partial a_k}q_l(\mathbf{a}) = 0. \qquad (7)$$

This is a linear, homogeneous partial differential equation in $q_l(\cdot)$, and can be solved via the method of characteristics (c.f. Courant, 1962, Chapter I.5 and II.2). The characteristic curve, i.e. the J-dimensional subspace on which $q_l(\mathbf{a})$ remains constant, can be obtained as follows. Parametrize $a_j = a_j(r)$, $j = 0, 1, \ldots J$ and consider

$$0 = \frac{dq_l}{dr} = \frac{\partial q_l(\mathbf{a})}{\partial a_l}\frac{da_l(r)}{dr} + \sum_{k=0,k\neq l}^{J}\frac{\partial q_l(\mathbf{a})}{\partial a_k}\times\frac{da_k(r)}{dr}.$$

Comparing with (7), we get

$$\frac{da_k(r)}{dr} = 1, k = 0, 1, ...J,$$

implying the so-called "characteristic" Ordinary Differential Equations:

$$\frac{da_k}{da_l} = 1, k = 0, ...l-1, l+1, ..., J, \quad (8)$$

with generic solutions $a_k - a_l = c_k$, $k = 0, ...l-1, l+1, ..., J$. This means that general solutions to (7) are of the form

$$q_l(\mathbf{a}) = H^l(a_0 - a_l, a_1 - a_l, ..., a_{l-1} - a_l, a_{l+1} - a_l, ...a_J - a_l), \quad (9)$$

where $H^l(\cdot)$ is any arbitrary continuously differentiable function. Thus $q_l(\mathbf{a})$ depends on the $(J+1)$-dimensional argument $(a_0, a_1, a_2, ...a_J)$ through a J-dimensional vector

$$(a_1 - a_l, a_2 - a_l, ..., a_{l-1} - a_l, a_{l+1} - a_l, ...a_J - a_l).$$

That (9) is a solution to (7) can also be verified directly by partially differentiating the RHS of (9), and verifying that it satisfies (7). Finally, note that

$$\begin{aligned} & (a_0 - a_l, a_1 - a_l, ..., a_{l-1} - a_l, a_{l+1} - a_l, ...a_J - a_l) \\ = \quad & (p_l, p_l - p_1, ..., p_l - p_{l-1}, p_l - p_{l+1}, ...p_l - p_J), \end{aligned}$$

and so (9) implies that $q_l(\mathbf{a})$ does not depend on income. Since l is arbitrary, we have shown that Slutsky symmetry implies that income effects are absent.

3 Rationalizability under Income-Effects

The previous section raises the question of whether utility maximization in a setting of multinomial choice that allows for income effects impose any restriction on choice-probabilities. In other words, is there a counterpart of Slutsky symmetry under income effects? In this section, we state that counterpart, and show that this analog, plus a set of shape-restrictions on choice-probabilities are together sufficient for rationalizability.

Counterpart of Slutsky Symmetry: Let there be $J+1$ exclusive and indivisible alternatives, indexed by $j = 0, 1, ..., J$. A consumer can choose one among these $J+1$ alternatives, plus a quantity z of a continuous numeraire that they can buy after paying for the indivisible good, subject to the budget constraint $z \leq y - p_j$, where y is the consumer's income, and p_j is the price of alternative j faced by the consumer. We assume preferences are non-satiated in the numeraire, and denote the amount of numeraire consumed upon having bought alternative j by $a_j = y - p_j$, with $a_0 = y$ corresponding to choosing the outside option 0. Denote the (structural) probability of choosing alternative $j \in \{0,...,J\}$ at $\mathbf{a} \equiv (a_0,..,a_J)$ by $q_j(\mathbf{a})$. In words, if we randomly sample individuals from the population, and offer the vector $\mathbf{a}$ to each sampled individual, then a fraction $q_j(\mathbf{a})$ will choose alternative j, in expectation. Then our counterpart of Slutsky symmetry is:

(A): For any $\mathbf{a}$, and any pair of alternatives $k \neq l$, the ratio $\frac{\partial}{\partial a_k} q_l(\mathbf{a}) / \frac{\partial}{\partial a_l} q_k(\mathbf{a})$ depends only on a_k and a_l.

Motivation: To see where this restriction comes from, consider the above setting of multinomial choice, and let the utility from consuming the jth alternative and a quantity z of the numeraire be given by $U(j,z) + \varepsilon_j$. The $\{\varepsilon_j\}$, which represent unobserved heterogeneity in preferences, are allowed to have any arbitrary and unspecified joint distribution in the population (subject to the resulting choice probability functions being smooth). If $U(j,\cdot)$ is strictly increasing (i.e. non-satiation in the numeraire), then we can replace $z = y - p_j \equiv a_j$, and rewrite the consumer problem as

$$\max_{j \in \{0,1,...,J\}} \left[U(j,a_j) + \varepsilon_j\right]. \tag{10}$$

To allow for income effects, we let $U(j,a_j) \equiv h_j(a_j)$, where $h_j(\cdot)$ are smooth, possibly nonlinear, strictly increasing, *unspecified* functions of the a_j's. When $h_j(\cdot)$ are nonlinear, the conditional choice-probabilities will depend on income, i.e., there are non-zero income effects. This structure is also observationally equivalent to a utility structure where unobserved heterogeneity is not additively separable from the a_j's (see below).

Now, for the above set-up, the choice probability for the 0th alternative is given by

$$\begin{aligned} & q_0(\mathbf{a}) \\ = \; & \Pr\left(\cap_{j\neq 0}\left\{h_0(a_0) + \varepsilon_0 > h_j(a_j) + \varepsilon_j\right\}\right) \\ = \; & \Pr\left[\cap_{j\neq 0}\left\{h_0(a_0) - h_j(a_j) > \varepsilon_j - \varepsilon_0\right\}\right] \\ = \; & \int_{-\infty}^{\infty}\int_{-\infty}^{(h_0(a_0)-h_1(a_1))+\varepsilon_0} \cdots \int_{-\infty}^{(h_0(a_0)-h_J(a_J))+\varepsilon_0} g(\varepsilon)\, d\varepsilon_J ... d\varepsilon_1 d\varepsilon_0. \end{aligned} \tag{11}$$

Therefore, by the first fundamental theorem of calculus,

$$\begin{aligned} & \frac{\partial}{\partial a_1} q_0(\mathbf{a}) \\ = \; & -h_1'(a_1)\left[\begin{array}{c} \int_{-\infty}^{\infty}\int_{-\infty}^{\varepsilon_0 + h_0(a_0) - h_2(a_2)} \cdots \int_{-\infty}^{\varepsilon_0 + h_0(a_0) - h_J(a_J)} g\left(\begin{array}{c} \varepsilon_0, \\ (h_0(a_0) - h_1(a_1)) + \varepsilon_0, \\ \varepsilon_2, ...\varepsilon_J \end{array}\right) \\ d\varepsilon_J ... d\varepsilon_2 d\varepsilon_0 \end{array}\right]. \end{aligned} \tag{12}$$

Similarly,

$$q_1(\mathbf{a}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\varepsilon_1 + h_1(a_1) - h_0(a_0)} \ldots \int_{-\infty}^{\varepsilon_1 + h_1(a_1) - h_J(a_J)} g(\varepsilon)\, d\varepsilon_J \ldots d\varepsilon_2 d\varepsilon_0 d\varepsilon_1,$$

implying by the first fundamental theorem and chain-rule that

$$\begin{aligned}
& \frac{\partial}{\partial a_0} q_1(\mathbf{a}) \\
= \; & -h_0'(a_0) \int_{-\infty}^{\infty} \int_{-\infty}^{h_1(a_1) - h_2(a_2) + \varepsilon_1} \ldots \int_{-\infty}^{h_1(a_1) - h_J(a_J) + \varepsilon_1} g\begin{pmatrix} h_1(a_1) - h_0(a_0) + \varepsilon_1, \\ \varepsilon_1, \varepsilon_2, \ldots \varepsilon_J \end{pmatrix} d\varepsilon_J \ldots d\varepsilon_2 d\varepsilon_1 \qquad (13) \\
\overset{(1)}{=} \; & -h_0'(a_0) \int_{-\infty}^{\infty} \int_{-\infty}^{s_0 + h_0(a_0) - h_2(a_2)} \ldots \int_{-\infty}^{s_0 + h_0(a_0) - h_J(a_J)} g\begin{pmatrix} s_0, \\ s_0 - h_1(a_1) + h_0(a_0), \\ \varepsilon_2, \ldots \varepsilon_J \end{pmatrix} \\
& \qquad d\varepsilon_J \ldots d\varepsilon_2 ds_0 \\
= \; & \frac{h_0'(a_0)}{h_1'(a_1)} \frac{\partial}{\partial a_1} q_0(\mathbf{a}), \text{ using (12)},
\end{aligned}$$

where the second equality $\overset{(1)}{=}$ follows by substituting $s_0 = h_1(a_1) - h_0(a_0) + \varepsilon_1$ in (13).

The same argument can be repeated for any other pair of alternatives $l \neq k$, to obtain

$$\frac{\frac{\partial}{\partial a_k} q_l(\mathbf{a})}{\frac{\partial}{\partial a_l} q_k(\mathbf{a})} = \frac{h_k'(a_k)}{h_l'(a_l)}, \qquad (14)$$

for all $\mathbf{a}$, and it is clear that the RHS of (14) depends only on a_k and a_l, and thus satisfies condition (A) above. As an aside, note that for the RHS of (14) to be identically equal to 1 (the Daly-Zachary condition), we must have that $h_l(a_l) = \beta_0 + \beta_1 a_l$ for some β_0, β_1. To see this, first note that when evaluated at $a_k = a_l = c$, condition (14) yields $\frac{h_k'(c)}{h_l'(c)} = 1$ for all c, implying $h_k(c) = h_l(c) + k$ for all c. Using this, we have that

$$1 = \frac{h_k'(a_k)}{h_l'(a_l)} = \frac{h_l'(a_k)}{h_l'(a_l)} \Rightarrow h_l''(a) = 0,$$

implying $h_l(a_l) = \beta_0 + \beta_1 a_l$, and thus the choice-probabilities cannot display income-effects.

Remark 1 *Condition (14) has no relation with the Independence of Irrelevant Alternatives (IIA) property. Indeed, the model above will* ***not*** *have the IIA property if the ε_js are correlated across alternatives (i.e. across j), but it will continue to satisfy (14), since uncorrelatedness of εs was not used to derive (14).*

Main Result: We are now ready to state and prove our main result. The result is that the counterpart of Slutsky symmetry stated above, plus two shape-restrictions on $q_j(\cdot)$'s are jointly *sufficient* for rationalizability, i.e., under those restrictions on $q_j(\cdot)$'s, we can find a set of utility functions and a joint distribution of unobserved preference heterogeneity, such that individual maximization of these utilities will indeed produce the conditional choice-probabilities $\{q_j(\cdot)\}$, $j = 0,1,...,J$.

To state and prove this result, we will use the following additional notation: let $\mathbf{a}_{-j}$ denote the vector $(a_0, a_1, ...a_{j-1}, a_{j+1}, ...a_J)$ and let $\lim_{\mathbf{a}_{-j}\downarrow-\infty}$ denote that each component of $\mathbf{a}_{-j}$ goes to $-\infty$.

Theorem 1 *Suppose that the following three conditions are satisfied by the choice-probabilities $\{q_j(\mathbf{a})\}$:*

(i) For each $j = 0,1,...,J$, and each $\mathbf{a}$, $q_j(\mathbf{a})$ is strictly increasing in a_j and strictly decreasing in a_k for $k \neq j$, continuously differentiable in each argument, and for all j, $\lim_{\mathbf{a}_{-j}\downarrow-\infty} q_j(\mathbf{a}) = 1$ and $\lim_{a_j\downarrow-\infty} q_j(\mathbf{a}) = 0 = 1 - \lim_{a_j\uparrow\infty} q_j(\mathbf{a})$;

(ii) given an alternative m and any other alternatives $j \neq m$ and any $\mathbf{a}$ satisfying $\frac{\partial}{\partial a_j} q_m(\mathbf{a}) \neq 0$, the ratio $\frac{\partial}{\partial a_m} q_j(\mathbf{a}) / \frac{\partial}{\partial a_j} q_m(\mathbf{a})$ does not depend on a_k, for $k \notin \{m, j\}$, and has uniformly bounded derivatives with respect to a_m and a_j;

(iii) for each $r = 0,1,...J$, the Jth order cross partial derivatives $\frac{\partial^J}{\partial a_0 \partial a_1 ... \partial a_{r-1} \partial a_{r+1} ... \partial a_J} q_r(\mathbf{a})$ exist, are continuous, and satisfy $(-1)^J \frac{\partial^J}{\partial a_0 \partial a_1 ... \partial a_{r-1} \partial a_{r+1} ... \partial a_J} q_r(\mathbf{a}) \geq 0$.

Then there exist random variables $\mathbf{V} = (V_0, V_1, ..., V_{m-1}, V_{m+1}, ..., V_J)$ with support $\mathscr{V} \sqsubseteq R^J$, and functions $w_j(a, v_j) : \mathbb{R} \times \mathscr{V}_j \to \mathbb{R}$, such that $w_j(\cdot, v_j)$ are strictly increasing and continuous, $w_m(a_m, v_m) \equiv a_m$, and

$$q_j(a_0, a_1, ..., a_J) = \int_{\mathscr{V}} \cap_{k\neq j} 1\{w_j(a, v_j) \geq w_k(a_k, v_k)\} f(\mathbf{v}) d\mathbf{v}$$

here $f(\cdot)$ denotes the joint density function of $\mathbf{V}$ on $\mathscr{V}$. Thus the utility functions $\{w_j(a, v_j)\}$ and heterogeneity distribution $f(\cdot)$ rationalize the choice probabilities $\{q_j(\mathbf{a})\}$.

Condition (i) is intuitive, and corresponds to preferences being non-satiated in the quantity of numeraire. Indeed, if choice probabilities are generated by the structure

$$q_j(\mathbf{a}) = \int_{\mathscr{V}} 1\left\{W_j(a_j, \eta) \geq \max_{r\in\{0,1,...J\}\setminus\{j\}} W_r(a_r, \eta)\right\} f(\eta) d\eta,$$

where $W_j(, \eta)$ are strictly increasing and continuous, and their distributions sufficiently smooth, then condition (i) must hold. Condition (iii) is related to the existence of a density function for unobserved heterogeneity. For models with *parametrically specified* heterogeneity distributions, condition (iii) was previously used to recover underlying utility functions (c.f. McFadden, 1978, just above Eqn. 12). The motivation for condition (ii) was discussed right before Theorem 1.

Proof. WLOG take $m = 0$, and use condition (ii) of the theorem to define

$$t_{j0}(a_j, a_0) \equiv \frac{\partial}{\partial a_0} q_j(\mathbf{a}) / \frac{\partial}{\partial a_j} q_0(\mathbf{a}) \geq 0. \tag{15}$$

Now, because $\sum_{j=0}^{J} q_j(\mathbf{a}) = 1$, differentiating both sides w.r.t. a_0 gives

$$\frac{\partial}{\partial a_0} q_0(\mathbf{a}) + \sum_{j=1}^{J} \frac{\partial}{\partial a_0} q_j(\mathbf{a}) = 0. \tag{16}$$

Substituting (15) in (16), we get the linear, homogeneous, partial differential equation in $q_0(\cdot)$:

$$\frac{\partial}{\partial a_0} q_0(\mathbf{a}) + \sum_{j=1}^{J} \frac{\partial}{\partial a_j} q_0(\mathbf{a}) \times t_{j0}(a_j, a_0) = 0. \tag{17}$$

This PDE can be solved via the method of characteristics (c.f. Courant, 1962, Chapter I.5 and II.2). The characteristic curve, i.e. the J-dimensional subspace on which $q_0(a)$ remains constant, can be obtained as follows. Parametrize $a_j = a_j(r)$ $j = 0, 1, ...J$ and consider

$$0 = \frac{dq_0}{dr} = \frac{\partial q_0(\mathbf{a})}{\partial a_0} \frac{da_0(r)}{dr} + \sum_{j=1}^{J} \frac{\partial q_0(\mathbf{a})}{\partial a_j} \times \frac{da_j(r)}{dr}.$$

Comparing with (17), we get

$$\frac{da_0(r)}{dr} = 1,\ \frac{da_j(r)}{dr} = t_{j0}(a_j, a_0),\ j = 1, ...J,$$

implying the characteristic ordinary differential equations:

$$\frac{da_j}{da_0} = t_{j0}(a_j, a_0), \tag{18}$$

for $j = 1, ..., J$. Using Picard's theorem and the principle of solving linear homogeneous PDEs (see Appendix), we obtain the general solutions of (18) given by $\omega_j(a_j, a_0) = cons$, where $\omega_j(a_j, a_0)$ is differentiable, strictly increasing in a_0 and strictly decreasing in a_j, and satisfies

$$\frac{\partial \omega_j(a_j, a_0)}{\partial a_0} + \frac{\partial \omega_j(a_j, a_0)}{\partial a_j} t_{j0}(a_j, a_0) = 0, \tag{19}$$

and also, using (15)

$$-\frac{\partial \omega_j(a_j, a_0)}{\partial a_0} / \frac{\partial \omega_j(a_j, a_0)}{\partial a_j} \equiv \frac{\partial}{\partial a_0} q_j(\mathbf{a}) / \frac{\partial}{\partial a_j} q_0(\mathbf{a}). \tag{20}$$

A general solution $q_0(\mathbf{a})$ is therefore of the form

$$q_0(\mathbf{a}) = H_0(\omega_1(a_1, a_0), \omega_2(a_2, a_0), ..., \omega_J(a_J, a_0)), \tag{21}$$

where $H_0(\cdot)$ can be chosen to be strictly increasing and C^1 in each argument, and with continuous Jth order cross partial derivatives. In particular, any J dimensional continuously differentiable C.D.F. $H_0(\cdot)$ would produce an admissible solution. Since $q_0(\mathbf{a})$ is observed, the exact functional form of $H_0(\cdot)$ is pinned down by (21), for any set of solutions $\omega_j(\cdot,\cdot)$ to the ODEs (18). This corresponds to the so-called "initial condition" in the PDE nomenclature. In particular, given any a_0, the value of $H_0(x_1,x_2,...x_J)$ at any vector $(x_1,x_2,...x_J)$ is given by

$$H_0(x_1,x_2,...x_J) = q_0(a_0,b_1(x_1,a_0),...b_J(x_J,a_0)),$$

where $b_j(x_j,a_0)$ is defined by

$$\omega_j(b_j(x_j,a_0),a_0) = x_j \tag{22}$$

In this construction, the choice of a_0 is immaterial. That is, for two choices $a_0 \neq a_0'$,

$$\begin{aligned}
& q_0(a_0,b_1(x_1,a_0),...b_J(x_J,a_0)) \\
= \; & H_0(\omega_1(b_1(x_1,a_0),a_0),\omega_2(b_2(x_2,a_0),a_0),...,\omega_J(b_J(x_J,a_0),a_0)) \\
= \; & H_0(x_1,x_2,...x_J) \\
= \; & H_0\left(\omega_1\left(b_1\left(x_1,a_0'\right),a_0'\right),\omega_2\left(b_2\left(x_2,a_0'\right),a_0'\right),...,\omega_J\left(b_J\left(x_J,a_0'\right),a_0'\right)\right) \\
= \; & q_0\left(a_0',b_1\left(x_1,a_0'\right),...b_J\left(x_J,a_0'\right)\right).
\end{aligned}$$

Having obtained the $\omega_j(\cdot,\cdot)$'s from (18) and (19), for each $j = 1,...J$, define the function $w_j(a_j,v)$ by inversion, i.e.

$$w_j(a_j,v) = \left\{a_0 : \omega_j(a_j,a_0) = v\right\}. \tag{23}$$

Note that by construction, $w_j(a_j,v)$ is strictly increasing and continuous in a_j for each v. The $w_j(\cdot,\cdot)$'s will play the role of 'utilities' in our proof of integrability. Set $w_0(a_0,v_0) \equiv a_0$.

We now show how to construct the distribution of heterogeneity. Let $\bar{\mathscr{V}}_j$ denote the co-domain of $\omega_j(\cdot,\cdot)$, and let

$$\mathscr{V}_j = \bar{\mathscr{V}}_j \cap \left\{\omega_j(a_j,a_0) : \prod_{j=1}^{J}\left\{\frac{\partial}{\partial a_0}\omega_j(a_j,a_0) \times \frac{\partial}{\partial a_j}\omega_j(a_j,a_0)\right\} \neq 0\right\},$$

and let $\mathscr{V} \equiv \times_{j=1}^{J}\mathscr{V}_j$. Now, given any vector $\mathbf{v} \equiv (v_1,...,v_J) \in \mathscr{V}$, define the cumulative distribution function at $\mathbf{v}$ as

$$F(v_1,...,v_J) = q_0(a_0,a_1,...,a_J),$$

where the vector $(a_0,a_1,...,a_J)$ satisfies $v_j = \omega_j(a_j,a_0)$, for each $j = 1,...J$. It follows from (21) and (22) that this function is well-defined. The above CDF implies the density function $f : \mathscr{V} \to \mathrm{R}_+$:

$$
\begin{aligned}
& f(v_1,...,v_J) \\
= \quad & \frac{\frac{\partial^J}{\partial a_1 ... \partial a_J} q_0(a_0,a_1,...,a_J)\,|_{v_j=\omega_j(a_j,a_0),\ j=1,...J}}{\prod_{j=1}^J \frac{\partial}{\partial a_j}\omega_j(a_j,a_0)\,|_{v_j=\omega_j(a_j,a_0),\ j=1,...J}} \qquad (24) \\
= \quad & \frac{\frac{\partial^{J-1}}{\partial a_1 ... \partial a_{k-1}\partial a_{k+1}...\partial a_J}\frac{\partial}{\partial a_k} q_0(a_0,a_1,...,a_J)\,|_{v_j=\omega_j(a_j,a_0),\ j=1,...J}}{\prod_{j=1}^J \frac{\partial}{\partial a_j}\omega_j(a_j,a_0)\,|_{v_j=\omega_j(a_j,a_0),\ j=1,...J}}, \text{ for any } k \in \{1,...,J\} \\
= \quad & \frac{\frac{\partial^{J-1}}{\partial a_1 ... \partial a_{k-1}\partial a_{k+1}...\partial a_J}\left[-\underbrace{\frac{\frac{\partial}{\partial a_k}\omega_k(a_k,a_0)}{\frac{\partial \omega_k(a_k,a_0)}{\partial a_0}}}_{\text{does not depend on } a_1...a_{k-1},a_{k+1}...a_J} \times \frac{\partial}{\partial a_0} q_k(a_0,a_1,...,a_J)\right]|_{v_j=\omega_j(a_j,a_0)}}{\prod_{j=1}^J \frac{\partial}{\partial a_j}\omega_j(a_j,a_0)\,|_{v_j=\omega_j(a_j,a_0),\ j=1,...J}}, \text{ from (20)} \\
= \quad & -\frac{\frac{\partial^{J-1}}{\partial a_1 ... \partial a_{k-1}\partial a_{k+1}...\partial a_J}\frac{\partial}{\partial a_0} q_k(a_0,a_1,...,a_J)\,|_{v_j=\omega_j(a_j,a_0),\ j=1,...J}}{\frac{\frac{\partial}{\partial a_0}\omega_k(a_k,a_0)}{\frac{\partial}{\partial a_k}\omega_k(a_k,a_0)} \times \prod_{j=1}^J \frac{\partial}{\partial a_j}\omega_j(a_j,a_0)\,|_{v_j=\omega_j(a_j,a_0),\ j=1,...J}} \\
= \quad & -\frac{\frac{\partial^{J-1}}{\partial a_1 ... \partial a_{k-1}\partial a_{k+1}...\partial a_J}\frac{\partial}{\partial a_0} q_k(a_0,a_1,...,a_J)\,|_{v_j=\omega_j(a_j,a_0),\ j=1,...J}}{\frac{\partial}{\partial a_0}\omega_k(a_k,a_0) \times \prod_{j=1,j\neq k}^J \frac{\partial}{\partial a_j}\omega_j(a_j,a_0)\,|_{v_j=\omega_j(a_j,a_0),\ j=1,...J}}. \qquad (25)
\end{aligned}
$$

Since $\frac{\partial^J}{\partial a_0 \partial a_1 ... \partial a_{k-1}\partial a_{k+1}\partial a_J} q_k(a_0,a_1,...a_J)$ has sign $(-1)^J$ and $\frac{\partial}{\partial a_j}\omega_j(a_j,a_0) < 0$, and $\frac{\partial}{\partial a_0}\omega_j(a_j,a_0) > 0$ on $\mathscr{V}$, each of the above expressions has numerator and denominator of the same sign, and is thus non-negative. We verify below that this joint density integrates to 1.

We now show that the above construction of $w_j(\cdot,\cdot)$ (c.f. (23)) and the joint density of heterogeneity (24) and (25) will indeed produce the original choice probabilities. To see this for alternative 1, consider the integral

$$
\begin{aligned}
& \int_{\mathscr{V}} 1\left\{w_1(a_1,v_1) \geq \max_{k\in\{0,2,...J\}} w_k(a_k,v_k)\right\} f(v_1,v_2,...,v_1)\,dv_1...dv_J \\
= \quad & \int_{\mathscr{V}} 1\left[v_1 \geq \omega_1(a_1,a_0), \cap_{k\in\{2,...J\}} 1\{v_k \leq \omega_k(a_k, w_1(a_1,v_1))\}\right] f(v_1,v_2,...,v_1)\,dv_1...dv_J
\end{aligned}
$$

Consider the substitution $(v_1,v_2,...v_J) \to (x_1,x_2,...x_J)$ given by $v_1 = \omega_1(a_1,x_1)$ (so that $x_1 = w_1(a_1,v_1)$), and for $k = 2, ..., J$, $v_k = \omega_k(x_k, x_1)$, which transforms the above integral to

$$\begin{aligned}
& \int_{a_0}^{\infty}\int_{a_2}^{\infty}...\int_{a_J}^{\infty}\left[\begin{array}{c} f(\omega_1(a_1,x_1),\omega_2(x_2,x_1)...,\omega_J(x_J,x_1)) \\ \times\left|\frac{\partial\omega_1(a_1,x_1)}{\partial x_1}\times\prod_{k=2}^{J}\frac{\partial\omega_j(x_j,x_1)}{\partial x_j}\right| \end{array}\right]dx_J...dx_2dx_1 \\
= & \int_{a_0}^{\infty}\int_{a_2}^{\infty}...\int_{a_J}^{\infty}\left[\begin{array}{c} f(\omega_1(a_1,x_1),\omega_2(x_2,x_1)...,\omega_J(x_J,x_1)) \\ \times(-1)^{J-1}\times\frac{\partial\omega_1(a_1,x_1)}{\partial x_1}\times\prod_{k=2}^{J}\frac{\partial\omega_j(x_j,x_1)}{\partial x_j} \end{array}\right]dx_J...dx_2dx_1 \\
= & (-1)^{J-1}\times\int_{a_0}^{\infty}\int_{a_2}^{\infty}...\int_{a_J}^{\infty}\left\{-\frac{\partial^J}{\partial x_1\partial x_2...\partial x_J}q_1(x_1,a_1,x_2,...x_J)\right\}dx_J...dx_2dx_1, \text{ by (25)} \\
= & (-1)^{J}\times\int_{a_0}^{\infty}\int_{a_2}^{\infty}...\int_{a_J}^{\infty}\left\{\frac{\partial^J}{\partial x_1\partial x_2...\partial x_J}q_1(x_1,a_1,x_2,...x_J)\right\}dx_J...dx_2dx_1 \\
= & \int_{\infty}^{a_0}\int_{\infty}^{a_2}...\int_{\infty}^{a_J}\left\{\frac{\partial^J}{\partial x_1\partial x_2...\partial x_J}q_1(x_1,a_1,x_2,...x_J)\right\}dx_J...dx_2dx_1 \\
= & q_1(a_0,a_1,a_2,...a_J). \qquad (26)
\end{aligned}$$

Exactly analogous steps for $j = 2, ...J$, and using (25), lead to the conclusion that for all $j \geq 1$,

$$\begin{aligned}
& \int 1\left\{w_j(a_j,v_j) \geq \max_{k\in\{0,1,2,...J\}\backslash\{j\}} w_k(a_k,v_k)\right\} f(v_1,v_2,...,v_1)\,dv_1...dv_J \\
= & q_j(a_0,a_1,a_2,...a_J).
\end{aligned}$$

Also, note that

$$\begin{aligned}
& \int 1\left\{a_0 \geq \max_{k\in\{1,2,...J\}} w_k(a_k,v_k)\right\} f(v_1,v_2,...,v_J)\,dv_1...dv_J \\
= & \int_0^{\omega_1(a_1,a_0)}...\int_0^{\omega_J(a_J,a_0)} f(v_1,v_2,...,v_J)\,dv_J...dv_1 \\
\text{substitute } v_j \rightarrow & \; x_j \text{ satisfying } v_j = \omega_j(x_j,a_0) \\
= & \int_{a_1}^{\infty}\int_{a_2}^{\infty}...\int_{a_J}^{\infty} f(\omega_1(x_1,a_0),...,\omega_J(x_J,a_0))\left|\frac{\partial\omega_1(x_1,a_0)}{\partial x_1}...\frac{\partial\omega_J(x_J,a_0)}{\partial x_J}\right|dx_J...dx_1 \\
= & \int_{a_1}^{\infty}\int_{a_2}^{\infty}...\int_{a_J}^{\infty}(-1)^J\times f(\omega_1(x_1,a_0),...,\omega_J(x_J,a_0))\frac{\partial\omega_1(x_1,a_0)}{\partial x_1}...\frac{\partial\omega_J(x_J,a_0)}{\partial x_J}dx_J...dx_1 \\
= & \int_{\infty}^{a_1}...\int_{\infty}^{a_J}\frac{\partial^J}{\partial\alpha_1...\partial\alpha_J}q_0(a_0,\alpha_1,...\alpha_J)\,|_{\alpha_1=x_1,...\alpha_J=x_J}dx_J...dx_1, \text{ by (24)} \\
= & q_0(a_0,a_1,...a_J).
\end{aligned}$$

Finally, to show that the joint density (24) integrates to 1, use exactly the same substitution as the one leading to (26), and observe that

$$
\begin{aligned}
& \int f\left(v_1, v_2, \ldots, v_J\right) dv_1 \ldots dv_J \\
= & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[\begin{array}{c} f\left(\omega_1\left(a_1, x_1\right), \omega_2\left(x_2, x_1\right) \ldots, \omega_J\left(x_J, x_1\right)\right) \\ \times \left| \frac{\partial \omega_1\left(a_1, x_1\right)}{\partial x_1} \times \prod_{k=2}^{J} \frac{\partial \omega_j\left(x_j, x_1\right)}{\partial x_j} \right| \end{array} \right] dx_2 \ldots dx_J dx_1 \\
= & (-1)^J \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left\{ \frac{\partial^J}{\partial x_1 \partial x_2 \ldots \partial x_J} q_1\left(x_1, a_1, x_2, \ldots x_J\right) \right\} dx_2 \ldots dx_J dx_1 \\
= & q_1\left(-\infty, a_1, -\infty, \ldots - \infty\right) \\
= & 1,
\end{aligned}
$$

where $q_1\left(-\infty, a_1, -\infty, \ldots - \infty\right)$ denotes the limit of the choice probability of alternative 1 when the a's for all other alternatives are tend to $-\infty$.

Thus we have shown that a population endowed with our constructed $w_j\left(\cdot, v_j\right)$ as utilities, together with the joint density of heterogeneity given by (24) would indeed produce the choice probabilities $\left\{q_j\left(\cdot, \ldots \cdot\right)\right\}$ for each $j = 0, 1, \ldots J$. ■

Remark 2 *The* ***utility*** *function for each alternative* j*, viz.* $w_j\left(a_j, v_j\right)$*, constructed in the proof of Theorem 1, consists of a scalar heterogeneity* v_j*. However, the individual* ***demand*** *function for alternative* j *has* J *separate sources of heterogeneity, i.e.*

$$
\begin{aligned}
Q_j\left(\mathbf{a}, \mathbf{v}\right) & = 1\left\{ w_j\left(a_j, v_j\right) \geq \max_{r \in \{0,1,\ldots J\} \setminus \{j\}} w_r\left(a_r, v_r\right) \right\} \\
& = Q_j\left(a_0, a_1, \ldots a_J, \underbrace{v_1, v_2, \ldots, v_J}_{J \text{ dimensional heterogeneity}} \right)
\end{aligned}
$$

Thus, we have rationalized a $(J+1)$ *dimensional choice probability function via a J-dimensional heterogeneity distribution.*

4 Further Points

Identification: Theorem 1 can also be used to identify utilities and the heterogeneity distributions nonparametrically from choice-probabilities observed in a dataset. Nonparametric identification of multinomial choice models has been studied previously in the econometric literature (c.f. Matzkin, 1993). Since our proof of rationalizability presented in Theorem 1 is constructive, it provides an alternative and novel way to obtain identification by solving PDEs.

Toward that end, suppose that the choice-probabilities are generated by maximization of the utilities $u_j \equiv \{h_j(a_j) + \varepsilon_j\}$, $j = 0, ..., J$, where the $h_j(\cdot)$ functions are strictly increasing and continuous, and hence invertible. Observe that an observationally equivalent utility structure is where utility for the 0th alternative is a_0 and that for the jth alternative is $h_0^{-1}\left(h_j(a_j) + \underbrace{\varepsilon_j - \varepsilon_0}_{v_j}\right) \equiv w_j(a_j, v_j)$, in that these utilities will produce exactly the same choice probabilities as the $\{u_j\}$s. We work under this normalization from now on. We also note in passing that the $w_j(a_j, v_j)$ are not necessarily additive in the unobserved heterogeneity v_j.

Let $\mathbf{a}$ and $q_j(\mathbf{a})$ be as above. We can use the proof of Theorem 1 to identify the $w_j(a_j, v_j)$ functions and the joint distribution of $(v_1, ..., v_J)$ from the $\{q_j(\mathbf{a})\}$, as follows. First, note that

$$q_0(\mathbf{a}) = \Pr\left(\cap_{j\neq 0}\{a_0 > w_j(a_j, v_j)\}\right) = \Pr\left[\cap_{j\neq 0}\{v_j < \omega_j(a_j, a_0)\}\right],$$

so that

$$\frac{\partial}{\partial a_j} q_0(\mathbf{a}) = \frac{\partial}{\partial a_j}\omega_j(a_j, a_0) \times F_j(\omega_1(a_1, a_0), ..., \omega_J(a_J, a_0)), \tag{27}$$

where $F_j(\cdot)$ denotes the derivative of the joint distribution function of $\mathbf{v}$ w.r.t. its jth element. On the other hand,

$$\begin{aligned} q_j(\mathbf{a}) &= \Pr[w_j(a_j, v_j) > a_0, w_j(a_j, v_j) > w_1(a_1, v_1), ...w_j(a_j, v_j) > w_J(a_J, v_J)) \\ &= \Pr[v_j > \omega_j(a_j, a_0), v_1 < \omega_1(a_1, w_j(a_j, v_j)), ...v_J < \omega_J(a_J, w_j(a_j, v_j))) \\ &= \int_{\omega_j(a_j, a_0)}^{\infty}\int_{-\infty}^{\omega_1(a_1, w_j(a_j, v_j))} ... \int_{-\infty}^{\omega_J(a_J, w_j(a_j, v_j))} f(v_1, ..., v_J)\, dv_J ... dv_1 dv_j, \end{aligned}$$

and therefore, by chain-rule, the first fundamental theorem of calculus, and using $w_j(a_j, \omega_j(a_j, a_0)) = a_0$, we have that

$$\begin{aligned} \frac{\partial}{\partial a_0} q_j(\mathbf{a}) &= -\frac{\partial}{\partial a_0}\omega_j(a_j, a_0) \times \int_{-\infty}^{\omega_1(a_1, a_0)} ... \int_{-\infty}^{\omega_J(a_J, a_0)} f(v_1, ..., v_J)\, dv_J ... dv_1 \\ &= -\frac{\partial}{\partial a_0}\omega_j(a_j, a_0) \times F_j(\omega_1(a_1, a_0), ..., \omega_J(a_J, a_0)), \end{aligned} \tag{28}$$

and thus from (27) and (28), we have that

$$-\frac{\partial \omega_j(a_j, a_0)}{\partial a_0} / \frac{\partial \omega_j(a_j, a_0)}{\partial a_j} \equiv \frac{\partial}{\partial a_0} q_j(\mathbf{a}) / \frac{\partial}{\partial a_j} q_0(\mathbf{a}), \tag{29}$$

which is the same as (20). The RHS of (29) is observable from the data, and under the hypothesis of the model, is solely a function of a_0 and a_j, which is a testable implication. If this implication is not rejected, denote the RHS of (29) as $t_j(a_j, a_0)$ (this $t_j(\cdot, \cdot)$ can be estimated by, say a least squares projection of $\frac{\partial}{\partial a_0} q_j(\mathbf{a}) / \frac{\partial}{\partial a_j} q_0(\mathbf{a})$ on a polynomial sieve in a_j, a_0). Then solve the PDE

$$\frac{\partial \omega_j(a_j, a_0)}{\partial a_0} + \frac{\partial \omega_j(a_j, a_0)}{\partial a_j} t_{j0}(a_j, a_0) = 0,$$

for the $\omega_j(\cdot,\cdot)$'s as outlined above in (18), obtain the $w_j(a_j,v_j)$ by inverting the solution $\omega_j(a_j,a_0)$'s w.r.t. a_0, and the joint density of $\mathbf{v}$ using (24).

Incorporating Covariates: In our discussion above, choice probabilities $q_j(\cdot)$ defined in Section 2, correspond to so-called "structural" parameters in Econometrics. Estimating these from a non-experimental dataset might be non-trivial when observed budget sets (i.e. price and/or income) are correlated with unobserved individual preferences across the cross-section of consumers. A common empirical assumption is that budget sets and preferences are independent, conditional on a set of observed covariates. Hence it is useful to see how to adapt the above results to the presence of covariates.[3]

Suppose in addition to price and income, we also observe a vector of consumer characteristics, denoted by s, and a vector of characteristics z_j for each alternative $j=1,...,J$. Assume that the choice-probabilities are generated by maximization of the utilities $u_0 \equiv \{h_0(a_0,s)+\varepsilon_0\}$, and $u_j \equiv \{h_j(a_j,z_j,s)+\varepsilon_j\}$, $j=1,...,J$, where $h_0(a,s)$ and each $h_j(a,z,s)$ are strictly increasing and continuous in a, and hence invertible. Then an observationally equivalent utility structure is where utility for the 0th alternative is a_0 and that for the jth alternative is $h_0^{-1}\left(h_j(a_j,z_j,s)+\underbrace{\varepsilon_j-\varepsilon_0}_{v_j},s\right) \equiv w_j(a_j,z_j,v_j,s)$, which is in general not linear or separable in v_j. Working off this normalization, and essentially repeating the same steps as above holding z_j,s fixed, lead to the conclusion that for each z_j, s,

$$-\frac{\partial\omega_j(a_j,a_0,z_j,s)}{\partial a_0}\Big/\frac{\partial\omega_j(a_j,a_0,z_j,s)}{\partial a_j} \equiv \frac{\partial}{\partial a_0}q_j(\mathbf{a},\mathbf{z},s)\Big/\frac{\partial}{\partial a_j}q_0(\mathbf{a},\mathbf{z},s). \tag{30}$$

The RHS of (30) is observable from the data, and for each fixed z_j and s, is solely a function of a_0, a_j, which is a testable implication. If this implication is not rejected, denote the RHS of (30) as $t_j(a_j,a_0,z_j,s)$, just as above. Then for each each fixed z_j and s, solve the PDE

$$\frac{\partial\omega_j(a_j,a_0,z_j,s)}{\partial a_0}+\frac{\partial\omega_j(a_j,a_0,z_j,s)}{\partial a_j}t_j(a_j,a_0,z_j,s)=0,$$

to obtain the $\omega_j(a_j,a_0,z_j,s)$, invert w.r.t. a_0 to obtain the utilities $w_j(a_j,v_j,z_j,s)$ and the joint density of $\mathbf{v}$ using the analog of (24), where we utilize the inverse of $\omega_j(a_j,a_0,z_j,s)$ w.r.t. a_j, analogous to (22) above.

References

[3] If this *conditional* independence of preferences and budget sets is also suspect, then one needs to employ a "control function" type strategy (c.f. Blundell and Powell, 2004) to estimate the structural choice-probabilities. Indeed, our results above explore the connection between random utility models and "structural" choice probabilities. So, given the extensive econometric literature on estimating structural parameters under endogeneity, we refrain from discussing the consistent estimation of $q_j(\cdot)$ any further.

1. Anderson, S.P., De Palma, A. and Thisse, J.F., 1992. Discrete choice theory of product differentiation. MIT press.

2. Armstrong, M. and Vickers, J., 2015. Which demand systems can be generated by discrete choice?. Journal of Economic Theory, 158, pp.293-307.

3. Coddington EA. 1961. An introduction to ordinary differential equations. Dover Publishing, New York.

4. Courant, Richard, 1962. Methods of Mathematical Physics, Vol. 2, Interscience, New York.

5. Daly, A., and Zachary, S.,1978. Improved multiple choice models. In Hensher,D., Dalvi,Q.(Eds.), Identifying and Measuring the Determinants of Mode Choice, Teakfields, London.

6. Lewbel, A., 2001. Demand Systems with and without Errors. American Economic Review, 91(3), pp.611-618.

7. Mas-Colell, A., Whinston, M.D. and Green, J.R., 1995. Microeconomic theory. New York: Oxford university press.

8. Matzkin R., 1993. Nonparametric identification and estimation of polychotomous choice models. Journal of Econometrics, 58(1-2):137-68.

9. McFadden, D., 1973. Conditional logit analysis of qualitative choice behavior.

10. McFadden, D., 1978. Modeling the choice of residential location. Transportation Research Record, (673).

5 Appendix

Two basic ideas from the theory of partial and ordinary differential equations are used to prove Theorem 1. We will use the notation C^1 to indicate a function that is once continuously differentiable.

First consider the linear homogeneous PDE

$$\frac{\partial \sigma(x,y,z)}{\partial x} + g_2(x,y)\frac{\partial \sigma(x,y,z)}{\partial y} + g_3(x,z)\frac{\partial \sigma(x,y,z)}{\partial z} = 0. \tag{31}$$

Suppose g_2 and g_3 are C^1 and do not vanish simultaneously. Then a general solution to this equation is given by

$$\sigma(x,y,z) = \phi(h_2(x,y), h_3(x,z)),$$

where $\phi(\cdot)$ is *any* arbitrary C^1 function, and $h_2(x,y) = c_2$ and $h_3(x,z) = c_3$ are general solutions to the ordinary differential equations

$$\frac{dy}{dx} = g_2(x,y), \ \frac{dz}{dx} = g_3(x,z). \tag{32}$$

(See e.g. Courant, 1962, Chapter I.5, II.2). In particular, $\phi(\cdot,\cdot)$ can be chosen to be strictly increasing in both arguments.

The ODE (32) are referred to as the "characteristic equations" of the linear PDE (31), and existence of a solution to the PDE (31) amounts to existence of a solution of the ODE (32). The following lemma restates a global version of the Picard-Lindelöf theorem that establishes conditions for existence of a solution to a first-order ODE.

Lemma 1 (Picard-Lindelöf theorem) *Suppose that a function $g : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ is continuous, and on each strip $S_a = \{(x,y) : |x| \leq a, \ |y| < \infty\}$, $g(x,y)$ is Lipschitz in y. Then the ordinary differential equation $n'(x) = g(x, n(x))$, has a general solution $n(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$ with $n(\cdot)$ being C^1. (See, for instance, Coddington, 1961, Theorem 9 and corollary).*

This result is proved by showing that under the assumptions of the lemma, the map $n(\cdot) :\rightarrow \int_{x_0}^{x} g(s, n(s))\, ds$ for any arbitrary x_0 is a contraction, thereby ensuring, via the Banach fixed point theorem, the existence of $n(\cdot)$ satisfying

$$n(x) = n(x_0) + \int_{x_0}^{x} g(s, n(s))\, ds.$$