J. Econ. Theory 223 (2025) 105938

Contents lists available at ScienceDirect

Journal of Economic Theory

journal homepage: www.elsevier.com/locate/jet

Integrability and identification in multinomial choice models

Debopam Bhattacharya [1]

University of Cambridge, United Kingdom of Great Britain and Northern Ireland

A R T I C L E I N F O

JEL classification:
C14
C25
D11

Keywords:
Multinomial choice
Unobserved heterogeneity
Random utility
Integrability
Symmetry
Income effects
Partial differential equations
Nonparametric identification
Random coefficient models
Bounds on counterfactuals

A B S T R A C T

McFadden's random-utility model of multinomial choice has long been the workhorse of applied research. We establish shape-restrictions with respect to price and income which are necessary and sufficient for multinomial choice-probability functions to be rationalized via random-utility models with additive but nonparametric unobserved heterogeneity and general income-effects. Our proof is constructive, and facilitates nonparametric identification of preference-distributions without requiring identification-at-infinity type arguments. A corollary shows that symmetry, a key condition for previous rationalizability results, is equivalent to absence of income-effects. Our results imply theory-consistent nonparametric bounds for choice-probabilities on counterfactual budget-sets. They also apply to widely used random-coefficient models, upon conditioning on observable choice characteristics. The theory of partial differential equations plays a key role in our analysis.

1. Introduction

McFadden's random utility model of multinomial choice and its subsequent modifications (cf. Berry and Haile, 2021 and Gandhi and Nevo, 2021 for up-to-date surveys) have gained enormous popularity among applied economists. This paper is concerned with the micro-theoretic underpinning of such models, and in particular, on the question of 'integrability', i.e. which choice probability functions are logically consistent with a random utility model, in the tradition of Block and Marschak, 1960; Falmagne, 1978 and McFadden and Richter, 1990. Apart from obvious theoretical interest, this question has practical implications for empirical modeling of individual demand as well as predicting aggregate demand and welfare on counterfactual budget-sets that arise from a new tax or subsidy or changes in choice-sets due to addition or elimination of choice-options. In particular, any utility distribution that rationalizes a given demand dataset can be used, in addition to shape restrictions implied by economic theory, to construct nonparametric, theory-consistent bounds on such counterfactuals.

There has been comparatively more work on integrability in empirical demand models with *continuous* goods, cf. Lewbel, 2001. More recently, Dette et al., 2016 and Hausman and Newey, 2016 have derived integrability conditions for choice of a single continuous good. The multinomial discrete choice case differs fundamentally from the single continuous good setting because the price of different alternatives are generically distinct, unlike continuous choice where the per unit price is typically constant across choices.

E-mail address: debobhatta@gmail.com.

[1] The author acknowledges financial support from the European Research Council via a Consolidator Grant EDWEL, Project number 681565.

https://doi.org/10.1016/j.jet.2024.105938
Received 23 May 2023; Received in revised form 21 November 2024; Accepted 23 November 2024

Available online 29 November 2024
0022-0531/© 2024 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

In the present paper, we first show that in multinomial choice settings that allow for nonparametric unobserved heterogeneity and arbitrary income effects, there is a set of shape restrictions on conditional choice probability functions which together are sufficient for integrability. The proof of this result is constructive, and the rationalizing utility functions are obtained by inverting solutions of a system of linear, homogeneous partial differential equations (PDEs). The way in which PDEs arise here is unrelated to Roy's Identity (cf. Mas-Colell et al., 1995, Proposition 3.G.4); the partial derivatives appearing in the PDE are of the average *demand* function, not the indirect utility function. Together with an additional restriction, the above conditions are then shown to be both necessary and sufficient for the additive random utility model (ARUM) of McFadden. In our analysis of integrability, we leave the joint distribution of unobserved heterogeneity terms nonparametric. Unlike the computationally intensive algorithmic approach of McFadden and Richter, 1990, further investigated in Kitamura and Stoye, 2016 and Deb et al., 2021, our conditions are closed-form and analytical, and can therefore be imposed on choice probability functions during estimation; they are also global, in the sense that their forms do not depend on how many and which budget sets happen to be observed in a specific dataset. On the other hand, MR and KS's approach work under unrestricted heterogeneity, whereas our set-up is the model with additive heterogeneity but also covers more flexible models like the widely-used random coefficient setting (e.g. mixed logit) which, conditional on observed covariates, have an additive structure.

For binary choice with general (i.e. non-additive) heterogeneity, Bhattacharya, 2021 provided a set of necessary and sufficient conditions for integrability. For more general multinomial choice, Daly and Zachary, 1978 provided a set of closed-form, global conditions under which closed-form choice-probability functions can be justified as having arisen from preference maximization by a heterogeneous population. Konning and Ridder, 2003 formalized these results and provided further clarifications. These conditions were independently derived in Armstrong and Vickers, 2015, who improved upon the Daly-Zachary results by including an outside option in the choice set. In all of these results, a key condition for integrability is a symmetry condition on choice probabilities. As a corollary of our main result, we show that in the multinomial setting, Daly-Zachary's symmetry condition 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, but showing "sufficiency", i.e. that symmetry implies absence of income effects is non-trivial. Fosgerau et al., 2013 derived conditions for choice probabilities to be rationalizable by a random utility model that is additive in both unobserved heterogeneity *and income*, i.e. essentially assuming quasi-linear preferences. In particular, they consider a set-up where utilities are of the form $U_j = m_j + \varepsilon_j$, where m_j is the deterministic part of utility and ε_j is the researcher-unobserved idiosyncratic component, and derive conditions on the relationship between choice probabilities and the m_j's that are sufficient for integrability. Our integrability results may be viewed as a generalization of the Fosgerau et al. result to a set-up where utilities are of the form $U_j = h_j\left(m_j\right) + \varepsilon_j$ where $h_j(\cdot)$'s are unknown sub-utility functions. Further, we show how these integrability results can be used to nonparametrically identify the underlying preference distributions (i.e. the functions $h_j(\cdot)$'s and the joint distribution of the ε_j's) from the 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. This is in the spirit of existing results on identification of multinomial choice models, which rely on different assumptions, e.g. utilities being linearly separable in a covariate with large support, cf. Matzkin, 1993. More recently, Allen and Rehbeck, 2019 consider multinomial models with additive heterogeneity, and show how to identify the deterministic utility functions underlying the observed choice behavior; they do not separately show how to obtain the distribution of unobserved heterogeneity. More importantly, Allen and Rehbeck, 2019 does not deal with integrability issues, i.e. which choice probability functions are rationalizable by an additive random utility model, which is the main concern of the present paper. An important distinguishing feature of our set-up is that the arguments of choice-probability functions, viz. price and income, arise from budget constraints and they play important roles in the proof of integrability and the identification strategy via shape-restrictions on the choice probability functions. This is in contrast to mathematical/statistical approaches to identification of multinomial choice models (e.g. Matzkin, 1993; Lewbel, 2000; Fosgerau et al., 2013; Chernozhukov et al., 2019; Allen and Rehbeck, 2019) which treat the arguments of choice-probabilities in a more abstract way. From a purely methodological standpoint, achieving nonparametric identification by solving PDEs appears to be novel in the discrete choice literature.

Next, we discuss the empirical usefulness of our results by showing how they can be used (a) to analyze random coefficient models that are popular in applied work, e.g. McFadden-Train's mixed logit or the BLP model, and (b) to calculate theory-consistent bounds for demand and welfare on counterfactual budget sets, e.g. those resulting from prospective introduction of new taxes and subsidies, price-changes due to mergers and potential changes in choice sets e.g. due to removal of alternatives. We are not aware of any result in the literature that provides entirely nonparametric bounds for choice probabilities on counterfactual budget sets in multinomial settings.

The plan for the rest of the paper is as follows. Section 2 discusses integrability for multinomial choice in presence of arbitrary and unknown income effects, and presents Theorem 1, the key result of this paper, followed by a discussion of Daly-Zachary's symmetry condition and its connection with lack of income effects. Section 3 discusses four further points, viz. the implication of the integrability result for nonparametric identification of preference distributions, incorporation of covariates into the analysis, the applicability of these results to random coefficient models and using these results to calculate bounds on counterfactual choice probabilities. Section 4 concludes. A short appendix at the end presents two mathematical results on partial and ordinary differential equations that are used in this paper, as well as proofs of the main results.

2. Set-up and key results

Consider a setting of multinomial choice, where the discrete alternatives are indexed by $j = 0, 1, \ldots, J$, 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 $p_0 \equiv 0$. The support of price and income are assumed to be $\mathbb{R}_+ \times \mathbb{R}_+$. Let the utility from consuming the jth alternative and a quantity z of the numeraire[2] be given by $W_j(z, \varepsilon_j)$ where $W_j(\cdot, \varepsilon_j)$ is not necessarily linear, and its functional form is unknown to the analyst, and ε_j is unobserved heterogeneity in the consumer preferences. The consumer's problem is $\max_{j\in\{0,1,\ldots,J\},z} W_j(z, \varepsilon_j)$, subject to the budget constraint $z \le y - p_j$, where y is the consumer's income, p_j is the price of alternative j faced by the consumer. If $W_j(\cdot, \varepsilon_j)$ is strictly increasing (i.e. non-satiation in the numeraire), then we can rewrite the consumer problem as $\max_{j\in\{0,1,\ldots,J\}} W_j(a_j, \varepsilon_j)$, where $a_j \equiv y - p_j$, $a_0 = y$.[3] A leading subcase of our set-up is the Additive Random Utility Model (ARUM), where $W_j(a_j, \varepsilon_j) = h_j(a_j) + \varepsilon_j$, with the functions $h_j(\cdot)$ being strictly increasing.

Denote the 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. For clarity of exposition, our analysis suppresses covariates other than price and income, since 'integrability' results involve price and income effects on demand. One can add other covariates as additional arguments of choice probabilities in all our results. Also, we first express all our results in terms of $a_j \equiv y - p_j$, $a_0 = y$, instead of the standard p_j's and y. Here a_0 acts as the continuous numeraire. Subsequently, we re-state our results in terms of choice probabilities reparametrized as functions of prices and income. Lastly, throughout the paper, we will assume continuous differentiability of the choice probability function in prices and income to sufficient orders and, to avoid repetitions, not include this separately each time among the conditions for our results.

Main Result: The key question in this paper is what restrictions on choice-probabilities are imposed by utility maximization in the above setting of multinomial choice that allows for arbitrary and unknown income effects (corresponding to $W_j(\cdot, \varepsilon_j)$ being nonlinear in the first argument). We now state and prove our main result, viz. that two shape-restrictions and an invariance condition on the choice probability functions $\{q_j(\cdot)\}$, $j \in \{0, \ldots, J\}$ are necessary and sufficient for existence of a set of utility functions and a joint distribution of unobserved heterogeneity, such that individual maximization of these utilities will produce $\{q_j(\cdot)\}$, $j \in \{0, \ldots, J\}$ as the choice-probabilities. These three conditions are as follows:

Condition (C)

(i) For each $j = 0, 1, \ldots, J$, and each $\mathbf{a}$, $q_j(\mathbf{a})$ is strictly increasing in a_j and strictly decreasing in a_k for $k \neq j$ and continuously differentiable in each argument over $\mathbb{R}^{J+1}$;

(ii) $\frac{\partial}{\partial a_m} q_j(\mathbf{a}) / \frac{\partial}{\partial a_j} q_m(\mathbf{a})$ depends only on a_j, a_m, and is of the form

$$\frac{\partial}{\partial a_m} q_j(\mathbf{a}) / \frac{\partial}{\partial a_j} q_m(\mathbf{a}) = G_m(a_m) / G_j(a_j),$$

for some functions $G_j(\cdot), G_m(\cdot) > 0$, for all $j \neq m$.

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

Condition C(i), sometimes stated as good 0 being a gross substitute for every other good, corresponds to preferences being non-satiated in the quantity of numeraire. Indeed, if choice probabilities are generated by the structure

$$q_j(\mathbf{a}) = \int_{\mathcal{V}} 1\left\{ W_j(a_j, \varepsilon) \ge \max_{r\in\{0,1,\ldots J\}\setminus\{j\}} W_r(a_r, \varepsilon) \right\} dF(\varepsilon), \tag{1}$$

where $W_j(, \varepsilon)$ are strictly increasing and continuous, and their distributions sufficiently smooth (in particular, 'ties' occur with probability zero), then condition C(i) must hold. Condition C(i) is the key shape-restriction we work with for both integrability and identification results. For this, the meaning of the arguments of choice probabilities, viz. income and the different prices, are important, unlike in more abstract discrete choice models like Matzkin, 1993 and Allen and Rehbeck, 2019, who are concerned solely with identification, and therefore do not require arguments of the choice probability functions to have any substantive meaning or for the functions to satisfy shape restrictions.

Condition C(iii) is related to the existence of a non-negative probability density function for unobserved heterogeneity. The relationship will be clarified in the proof of Lemma 1 below. For models with *parametrically specified* heterogeneity distributions, condition (iii) was previously used to recover underlying utility functions (cf. McFadden, 1978 just above Eqn. 12, and McFadden, 1981).

Condition C(ii) represents the fundamental empirical content of a preference distribution characterized by *additive* unobserved heterogeneity (or one that is observationally equivalent to it). See the next subsection for how the additive heterogeneity model leads to this condition. Note that Condition C(ii) 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 (47), since uncorrelatedness of εs was not used to derive (47).

[2] It is natural to think of the numeraire as composite of all other goods, denominated in dollars. In particular, when a consumer chooses $j = 0$, i.e. the outside option, then she spends all her income on consumption of other goods.

[3] The support of the variables a_j are taken to be $\mathbb{R}$, as income minus price can be negative. In particular, an individual consumer's income can be low enough so that one or more options cost more than this level of income. For example, a luxury car or vacation package can cost well above a fast-food worker's salary.

Under smoothness, an empirical implication of C(ii) is that

$$\frac{\partial}{\partial a_k}\left\{\frac{\partial}{\partial a_m}q_j(\mathbf{a})/\frac{\partial}{\partial a_j}q_m(\mathbf{a})\right\}=0 \text{ for } k\neq j,m,$$

$$\frac{\partial^2}{\partial a_j\partial a_m}\left\{\ln\left(\frac{\partial}{\partial a_m}q_j(\mathbf{a})/\frac{\partial}{\partial a_j}q_m(\mathbf{a})\right)\right\}=0 \text{ for all } j\neq m.$$

The following result demonstrates that C(i)-C(iii) also constitute a *sufficient* condition for the choice probabilities to be generated by a preference structure with additive unobserved heterogeneity.

Theorem 1. *Assume that Conditions C(i), C(ii) and C(iii) hold. Then there exist strictly increasing, continuously differentiable utility functions* $h_j(\cdot):\mathbb{R}\to\mathbb{R}$*, and* J *dimensional unobserved heterogeneity* $(v_1,...,v_J)$ *with strictly positive joint density on a support* $\mathcal{V}\sqsubseteq\mathbb{R}^J$*, such that for all* $j=0,1,...,J$.

$$q_j(a_0,a_1,...,a_J)=\Pr\left[\cap_{k\neq j}1\left\{h_j(a_j)+v_j\geq h_k(a_k)+v_k\right\}\right]. \tag{2}$$

Conditions C(i), C(ii) and C(iii) are also necessary for (2) to hold (proof in Appendix).

The proof of Theorem 1 will utilize the following lemma, whose proof appears in the Appendix. This lemma and its proof will be referenced several times in the main text below. So we introduce it here for ease of reference.

Lemma 1. *Suppose conditions C(i) and C(iii) are satisfied by the choice-probabilities* $\{q_j(\mathbf{a})\}$*. Additionally, it holds that (ii') for any pair of 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*. Then there exist random variables* $\mathbf{V}=(V_0,V_1,...,V_{m-1},V_{m+1},...,V_J)$ *with support* $\mathcal{V}\sqsubseteq\mathbb{R}^J$ *and a joint distribution function* $F(\cdot)$*, and 'utility' functions* $w_j(a,v_j):\mathbb{R}\times\mathcal{V}_j\to\mathbb{R}$*, such that* $w_j(\cdot,v_j)$ *are strictly increasing and continuous,* $w_m(a_m,v_m)\equiv a_m$*, and*

$$q_j(a_0,a_1,...,a_J)=\int_{\mathcal{V}}\cap_{k\neq j}1\left\{w_j(a_j,v_j)\geq w_k(a_k,v_k)\right\}dF(\mathbf{v})$$

for each $j=0,1,...J$*. Thus the utility functions* $\{w_j(a,v_j)\}$ *and heterogeneity distribution* F *rationalize the choice probabilities* $\{q_j(\mathbf{a})\}$ *(proof in Appendix).*

Further Points: Two further points are worth noting in the context of Theorem 1. Firstly, note that an additive heterogeneity specification for utilities $W_j(a_j,\varepsilon)\equiv h_j(a_j)+\varepsilon_j$, which is shown below to be equivalent to Condition C(ii), implies an invariance of choice ordering over alternatives across consumers. For example, if in a budget situation (a), individual i chooses alternative 1 and individual k chooses alternative 2, then

$$\begin{bmatrix} h_1(a_1)+\varepsilon_{1i}>h_2(a_2)+\varepsilon_{2i},\ h_1(a_1)+\varepsilon_{1k}<h_2(a_2)+\varepsilon_{2k} \\ \Longrightarrow \varepsilon_{1i}-\varepsilon_{2i}>h_2(a_2)-h_1(a_1)>\varepsilon_{1k}-\varepsilon_{2k}. \end{bmatrix}.$$

Therefore, for any other budget set (b), it cannot be the case that

$$h_2(b_2)+\varepsilon_{2i}>h_1(b_1)+\varepsilon_{1i}.\text{and } h_1(b_1)+\varepsilon_{1k}>h_2(b_2)+\varepsilon_{2k}$$

i.e. in no budget situation can it be the case that individual i chooses alternative 2 and individual k chooses alternative 1. Via its equivalence with additive heterogeneity, condition C(ii) has therefore the same implication for choice. However, this implication pertains to two individuals choosing in multiple choice situations, and, as such, does not restrict behavior in a single cross-sectional setting.

Secondly, the *utility* function for each alternative j, viz. $h_j(a_j)+\varepsilon_j$, constructed in the proof of Theorem 1, consists of a scalar heterogeneity ε_j. However, the individual *demand* function for alternative j, denoted by $Q_j(\cdot,\cdot)$ has J separate sources of heterogeneity, i.e.

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

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

2.1. Motivation for C(ii) and Daly-Zachary's Symmetry

To see where condition C(ii) comes from, suppose the utility from consuming the jth alternative and a quantity z of the numeraire be given by $h_j(z) + \varepsilon_j$. 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). When $h_j(\cdot)$ are nonlinear, the conditional choice-probabilities will depend on income, i.e., there are non-zero income effects. Then it can be shown by a series of steps (see Appendix, proof of Theorem 1) that

$$\frac{\frac{\partial}{\partial a_k} q_l(\mathbf{a})}{\frac{\partial}{\partial a_l} q_k(\mathbf{a})} = \frac{h'_k(a_k)}{h'_l(a_l)} \tag{3}$$

Condition (3) was independently derived in Allen and Rehbeck, 2019 for the purpose of identifying the functions $h'_k(\cdot)$, when the additive structure for utilities $h_j(z)+\varepsilon_j$ is known to have produced the data. Daly-Zachary's symmetry conditions are simply a special case of (3), viz. that for any two alternatives $k, l \in \{0, 1, ..., J\}$, $k \neq l$,[4]

$$\frac{\partial}{\partial a_l} q_k(\mathbf{a}) = \frac{\partial}{\partial a_k} q_l(\mathbf{a}). \tag{4}$$

Indeed, the classic random utility model of multinomial choice (McFadden, 1981) assumes that the systematic part of the utility from consuming the jth alternative at income y and price p_j is given by

$$h_j(a_j) \equiv a_j, \tag{5}$$

where $a_j = y - p_j$ as above. Income effects are zero since demand depends on the a's via the differences $a_j - a_k = (y - p_j) - (y - p_k) = p_k - p_j$. Then (3) with $h_j(a_j) = a_j$, i.e. $h'_j(a_j) = 1$ implies

$$\frac{\frac{\partial}{\partial a_k} q_l(\mathbf{a})}{\frac{\partial}{\partial a_l} q_k(\mathbf{a})} = 1, \tag{6}$$

for all **a**. This shows that in the random utility model with no income effects, Daly-Zachary's symmetry condition holds. In fact, in the above set-up, symmetry implies absence of income effects. This can be shown simply as a special case of the steps, based on solving linear homogeneous PDEs, as outlined in the proof of Lemma 1 in the Appendix. In particular, (32) in the proof of Lemma 1 specializes in this case to

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

leading to the "characteristic" Ordinary Differential Equations:

$$\frac{da_k}{da_l} = 1, \; k = 0, ...l-1, l+1, ..., J, \tag{8}$$

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

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

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

$$\left(a_0 - a_l, a_1 - a_l, a_2 - a_l, ..., a_{l-1} - a_l, a_{l+1} - a_l, ...a_J - a_l\right).$$

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

$$\left(a_0 - a_l, a_1 - a_l, ..., a_{l-1} - a_l, a_{l+1} - a_l, ...a_J - a_l\right)$$

---

[4] Daly-Zachary defines choice probabilities as functions of price and income, $\bar{q}_j(p_0, p_1, ..., p_J, y)$. This is equivalent to our notation of $q_j(a_0, a_1, ...a_J)$ 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

$$q_j(a_0, a_1, ..., a_J) = \bar{q}_j(a_0 - a_1, a_0 - a_2, ..., a_0 - a_J), \text{ and}$$

$$\bar{q}_j(p_1, p_2, ..., p_J, y) = q_j(y, y - p_1, y - p_2, ..., y - p_J).$$

"Slutsky symmetry" in Daly-Zachary's notation is that $\partial\bar{q}_k/\partial p_j = \partial\bar{q}_j/\partial p_k$ for all $j \neq k$ (if alternative 0 is the outside option, then the corresponding condition is $\partial\bar{q}_0/\partial p_j = \partial\bar{q}_j/\partial y$), which is identical to (4) in our notation.

$$= \left(p_l, p_l - p_1, ..., p_l - p_{l-1}, p_l - p_{l+1}, ...p_l - p_J\right),$$

and so (9) implies that $q_l(\mathbf{a})$ does not depend on income. Since l is arbitrary, it follows that symmetry implies absence of income effects.

2.2. Conditions in standard form

Above, we have expressed choice probabilities and the conditions for our results in terms of the a_js, as opposed to p_js and y, since it is more natural to impose monotonicity of a function in its arguments, rather than on combination of derivatives with respect to arguments. If choice probabilities are instead expressed in the standard form with income and prices as arguments, one has

$$\begin{aligned} q_j\left(a_0, a_1, ..., a_J\right) &= \bar{q}_j\left(a_0, a_0 - a_1, ..., a_0 - a_J\right) \\ &= \bar{q}_j\left(y, p_1, ..., p_J\right) \equiv \bar{q}_j\left(y, \mathbf{p}\right). \end{aligned} \tag{10}$$

Then the shape restrictions, i.e. condition C(i) become: for each $j = 1, ...J$, $\partial\bar{q}_j(y,\mathbf{p})/\partial p_j \le 0$, $\partial\bar{q}_j(\mathbf{p},y)/\partial p_k \ge 0$ for all $k \ne j$, and $\sum_{k=1}^{J}\partial\bar{q}_j(y,\mathbf{p})/\partial p_k + \partial\bar{q}_j(y,\mathbf{p})/\partial y \le 0$ for all $j = 1, ...J$. The forms of these expressions bear similarity to Slutsky inequality conditions in standard, deterministic demand analysis for continuous goods. An important difference with the standard continuous case is that our condition is

$$\sum_{k=1}^{J}\partial\bar{q}_j(y,\mathbf{p})/\partial p_k + \partial\bar{q}_j(y,\mathbf{p})/\partial y \le 0, \tag{11}$$

in contrast to the standard case with J continuous goods, where the Slutsky negativity condition (cf. Varian, 1992 page 121) is

$$\sum_{k=1}^{J}\partial\bar{q}_j(y,\mathbf{p})/\partial p_k + \bar{q}_j(y,\mathbf{p}) \times \partial\bar{q}_j(y,\mathbf{p})/\partial y \le 0. \tag{12}$$

The inequality (11) can be given an axiom of revealed preference type interpretation. To see this easily, let $J = 1$ (i.e. a situation of binary choice), and note that (11) is equivalent to

$$\frac{\partial}{\partial a}\bar{q}_1\left(y + a, p_1 + a\right) \le 0 \tag{13}$$

for all a. This means that if both price and income rise by the same amount, then the fraction of consumers choosing 1 cannot increase. This is because by choosing 1, every consumer would get exactly the same utility as buying 1 in the initial situation, but because they have more income, choosing the 0 (not buying) option can give some consumers higher utility than when income was lower. The same idea applies to the case where $J \ge 2$.

As for Condition C(ii), note from (10) that

$$\frac{\partial q_j\left(a_0, a_1, ..., a_J\right)}{\partial a_0} = \sum_{k=1}^{J}\partial\bar{q}_j(y,\mathbf{p})/\partial p_k + \partial\bar{q}_j(y,\mathbf{p})/\partial y \tag{14}$$

$$\frac{\partial q_j\left(a_0, a_1, ..., a_J\right)}{\partial a_k} = \partial\bar{q}_j(y,\mathbf{p})/\partial p_k \text{ for all } k \ne 0. \tag{15}$$

Therefore, Condition C(ii) becomes: (A) for all $j = 1, 2, ..., J$,

$$\frac{\sum_{k=1}^{J}\partial\bar{q}_j(y,\mathbf{p})/\partial p_k + \partial\bar{q}_j(y,\mathbf{p})/\partial y}{\partial\bar{q}_0(y,\mathbf{p})/\partial p_j}$$

depends on $(y,\mathbf{p})$ only via $\left(y, y - p_j\right)$, i.e. via $\left(y, p_j\right)$; and (B) for all $j, k = 1, 2, ..., J$ with $j \ne k$, $\frac{\partial\bar{q}_j(y,\mathbf{p})/\partial p_k}{\partial\bar{q}_k(y,\mathbf{p})/\partial p_j}$ depends on $(y,\mathbf{p})$ only via $\left(y - p_k, y - p_j\right)$.

Finally, condition C(iii) is: for all $r = 1, 2, ..., J$,

$$\sum_{k=1}^{J}\frac{\partial}{\partial p_k}\left[\frac{\partial^{J-1}}{\partial p_1 ...\partial p_{r-1}\partial p_{r+1} ...\partial p_J}\bar{q}_r(y,\mathbf{p})\right] \ge 0.$$

3. Implications and further points

3.1. Identification

Theorem 1 can be used to identify utilities and the heterogeneity distributions nonparametrically from choice-probabilities observed in a dataset. Nonparametric identification of multinomial choice models (without any discussion of integrability) has been studied previously in the econometric literature, cf. Matzkin, 1993, 2007. Since our proof of integrability presented in Theorem 1 is constructive, it provides an alternative and novel way to obtain identification by solving PDEs. Unlike Matzkin, 1993, our identification strategy does not rely on identification-at-infinity type arguments nor on linear separability in a regressor with large support (cf. Matzkin, 2007), and instead requires smoothness. Allen and Rehbeck, 2019 provided an alternative way to prove identification using results from convex analysis.

Remark 1. We emphasize here that the integrability results presented above are the key contributions of our paper. The identification analysis is presented as a corollary thereof, and shows that one can analyze the two problems in a theoretically unified way.

Specifically, our identification approach is as follows. 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 utility functions $h_j(\cdot)$ are strictly increasing and continuous and hence invertible, but otherwise unknown and the joint distribution of the ε's is unknown. 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 (note that the $w_j(a_j, v_j)$ are not necessarily additive in the unobserved heterogeneity v_j). In other words, we have to normalize the utilities $\{h_j(a_j) + \varepsilon_j\}$, $j = 0, ..., J$ for identification. Here, we set $w_0(a_0, v_0) \equiv a_0$ and $w_j(a_j, v_j)$ to be strictly increasing in the numeraire a_j and invertible in the v_j for all $j \neq 0$. The goal is to identify the functional forms $w_j(a_j, v_j)$ and the joint distribution of the v_j's. We continue to assume that the functions $q_j(\mathbf{a})$ are smooth (have derivatives of required orders), and that the joint distribution of the v_j's admit a positive density on a connected support.

Let $\mathbf{a}$ and $q_j(\mathbf{a})$ be as above. We can 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

$$q_0(\mathbf{a}) = \Pr\left(\cap_{j\neq 0}\{a_0 > w_j(a_j, v_j)\}\right) = \Pr\left[\cap_{j\neq 0}\{v_j < \omega_j(a_j, a_0)\}\right], \tag{16}$$

where $\omega_j(a_j, a_0)$ denotes the inverse of $w_j(a_j, v_j)$ w.r.t. v_j at a_0 (i.e. $w_j(a_j, \omega_j(a_j, a_0)) \equiv a_0$). Therefore,

$$\frac{\partial}{\partial a_j} q_0(\mathbf{a}) = \frac{\partial}{\partial a_j}\omega_j(a_j, a_0) \times F_j(\omega_1(a_1, a_0), ..., \omega_J(a_J, a_0)), \tag{17}$$

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,

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

and therefore, by the 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

$$\begin{aligned} \frac{\partial}{\partial a_0} q_j(\mathbf{a}) &= -\frac{\partial}{\partial a_0}\omega_j(a_j, a_0) \times \int_{-\infty}^{\omega_1(a_1, a_0)} \cdots \int_{-\infty}^{\omega_J(a_J, a_0)} f(v_1, ..., v_J)\, dv_J ... dv_1 \\ &= -\frac{\partial}{\partial a_0}\omega_j(a_j, a_0) \times F_j(\omega_1(a_1, a_0), ..., \omega_J(a_J, a_0)), \end{aligned} \tag{18}$$

and thus from (17) and (18), we have that

$$-\frac{\partial \omega_j(a_j, a_0)}{\partial a_0} / \frac{\partial \omega_j(a_j, a_0)}{\partial a_j} \equiv \frac{\partial}{\partial a_0} q_j(\mathbf{a}) / \frac{\partial}{\partial a_j} q_0(\mathbf{a}). \tag{19}$$

The RHS of (19) is nonparametrically identifiable 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.[5] If this implication is not rejected, denote the RHS of (19) as $t_j(a_j, a_0)$. Then solve the PDE

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

[5] Allan and Rehbeck, 2019 note that this provides overidentifying restrictions for utility indices.

uniquely for the $\omega_j(\cdot,\cdot)$'s as outlined in the proof of Lemma 1 in the Appendix (see in particular (33) and (34)), where $\omega_j(a_j, a_0)$ is strictly increasing in a_0 and strictly decreasing in a_j, and 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 distribution of $\mathbf{v}$ using (16). Thus the values of $w_j(a_j, v_j)$ and the joint distribution of $\{v_j\}$'s are point-identified at those $\{v_j\}$s that equal $\omega_j(a_j, a_0)$ for some a_j, a_0 in the support of the a_j's. In particular, if the support $\mathcal{A}$ of the a_j's is rich enough that $q_0(\mathbf{a})$ takes all values in $[0,1]$ on $\mathcal{A}$, we can define the support of the v's as

$$\mathcal{V} = \left\{\omega_1(a_1, a_0), ..., \omega(a_J, a_0) : (a_0, a_1, ...a_J) \in \mathcal{A}\right\}$$

and by (16), the values of $w_j(a_j, v_j)$ and the joint distribution of $\{v_j\}$'s are point-identified on all of $\mathcal{V}$.

The route to identification outlined here can be challenging to implement fully nonparametrically in a finite dataset, as it involves estimating partial derivatives and numerically solving PDEs. This may be viewed as the cost of eschewing alternative routes to identification such as utilizing a special regressor with large support and/or using identification-at-infinity type methods.

3.2. Incorporating covariates

In our discussion above, choice probabilities $q_j(\cdot)$ defined in Section 2, correspond to so-called "average structural function", cf. Blundell and Powell, 2003, in the sense that the integral in (1) is with respect to the *marginal* distribution of unobserved heterogeneity, rather than the distribution conditional on price and income. 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.

Suppose in addition to price and income, we also observe a vector of characteristics z_j for each alternative $j = 1, ..., J$, and a set of individual specific covariates x. Assume that the choice-probabilities are generated by maximization of the utilities

$$u_0 \equiv \left\{h_0(a_0, x) + \varepsilon_0\right\}, \quad u_j \equiv \left\{h_j(a_j, z_j, x) + \varepsilon_j\right\}, j = 1, ..., J, \tag{20}$$

where $h_0(a, x)$ and each $h_j(a, z_j, x)$ 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

$$h_0^{-1}\left(h_j(a_j, z_j, x) + \underbrace{\varepsilon_j - \varepsilon_0}_{v_j}, x\right) \equiv w_j(a_j, z_j, x, v_j), \tag{21}$$

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 and x fixed, lead to the conclusion that for each z_j and x,

$$-\frac{\partial\omega_j(a_j, a_0, z_j, x)}{\partial a_0} \Big/ \frac{\partial\omega_j(a_j, a_0, z_j, x)}{\partial a_j} \equiv \frac{\partial}{\partial a_0} q_j(\mathbf{a}, \mathbf{z}, x) \Big/ \frac{\partial}{\partial a_j} q_0(\mathbf{a}, \mathbf{z}, x). \tag{22}$$

The RHS of (22) is observable from the data, and for each fixed z_j, x, is solely a function of a_0, a_j, which is a testable implication. If this implication is not rejected, denote the RHS of (22) as $t_j(a_j, a_0, z_j, x)$, just as above. Then for each fixed z_j, solve the PDE

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

to obtain the $\omega_j(a_j, a_0, z_j, x)$, invert w.r.t. a_0 to obtain the utilities $w_j(a_j, v_j, z_j, x)$ and the joint distribution of $\mathbf{v}$ using the analog of (38), where we utilize the inverse of $\omega_j(a_j, a_0, z_j, x)$ w.r.t. a_j, analogous to (37). If even *conditional* on covariates, independence of preferences and budget sets is suspect, then one needs to employ either a 'control function' type strategy (cf. Blundell and Powell, 2004; Imbens and Newey, 2009) or impose a 'special regressor' type structure (Lewbel, 2000) to estimate the average structural functions which, in our context, are the structural choice-probabilities $q_j(\cdot)$ defined in (1).

3.3. Empirical implications: bounds on counterfactuals

A key empirical implication of our results is that they can be used to obtain bounds for predicted demand on counterfactual budget sets. We demonstrate how to construct such bounds in the two leading cases of interest, viz. price changes and elimination/addition of alternatives.

Price Changes: Denote the support of observed price and income in the data by $\mathcal{A}$ (a discrete set), and suppose we have to *predict* demand for alternative 1 at a counterfactual $\mathbf{a}' = (a_0', a_1' ..., a_J') \notin \mathcal{A}$. Such counterfactual budget sets may arise due to potential price changes, e.g. those caused by taxes and subsidies or firm-mergers. To predict this counterfactual demand, let $\mathcal{A}_j$ denote the set of values of a_j's that appear in $\mathcal{A}$, and $\mathcal{A}_{jk}$ denote the collection of values taken by the pairs $\{a_j, a_k\}$, $j \neq k$ that appear in $\mathcal{A}$. Condition (i) of Theorem 1 yields the following bounds:

$$LB_j(\mathbf{a}') = \sup_{\mathbf{a}\in\mathcal{A}} \left\{ q_j(\mathbf{a}) : \ a'_j \le a_j, \ a_k \le a'_k \text{ for all } k \ne j \right\}$$
$$UB_j(\mathbf{a}') = \inf_{\mathbf{a}\in\mathcal{A}} \left\{ q_j(\mathbf{a}) : \ a'_j \ge a_j, \ a_k \ge a'_k \text{ for all } k \ne j \right\}. \tag{23}$$

These bounds in turn will provide bounds for welfare calculations corresponding to changes in prices (cf. Bhattacharya, 2018). The lower bound in (23) can be calculated by (i) selecting those observations for which $a'_j \le a_j$ and $a_k \le a'_k$ for all $k \ne j$, a one-line command in STATA, (ii) then calculating (nonparametrically) the choice probabilities at the selected a's, and (iii) picking the largest of them. Analogously for the upper bound. Manski, 2007 derives bounds on counterfactual choice probabilities using McFadden-Richter, 1990 type inequalities, while Tebaldi et al., 2023 provide an empirical application of such bounds under the assumption of quasilinear utilities.

To get simultaneous bounds on $\{q_j(\mathbf{a}')\}$, $j = 0, \ldots, J$, we have to impose the constraint that the sum of lower bounds and the sum of upper bounds over $j = 0, 1, \ldots J$ must equal 1. This amounts to finding the set of $\tilde{q}_j(\mathbf{a}')$, $j = 0, \ldots, J$ such that

$$LB_j(\mathbf{a}') \le \tilde{q}_j(\mathbf{a}') \le UB_j(\mathbf{a}'), \ \sum_{j=0}^{J} \tilde{q}_j(\mathbf{a}') = 1, \tag{24}$$

where $LB_j(\mathbf{a}')$ and $UB_j(\mathbf{a}')$, defined in (23), are point-identified and satisfy the shape restrictions of Theorem 1 (i). Note that (24) is a set of linear equality/inequality constraints in $\tilde{q}_j(\mathbf{a}')$ and can be computed using simplex methods. Molinari, 2020 discusses several substantive econometric problems that have such linear structure. The bounds (24) on demand in turn provide bounds for money-metric welfare effects corresponding to changes in prices of the products, or addition and elimination of options, since welfare expressions for such cases e.g. the average compensating variation, are closed-form functionals of choice probabilities, cf. Bhattacharya, 2018. The bounds in (24) are sharp because the choice probabilities $\{q_j(\mathbf{a}) \cup \tilde{q}_j(\mathbf{a}')\}_{j=0,\ldots,J}$ on $\mathcal{A} \cup \{\mathbf{a}'\}$ where $\tilde{q}_j(\mathbf{a}')$ satisfies (24), satisfy the shape-restrictions of Lemma 1, and can therefore be rationalized by the same utility functions and heterogeneity distribution as those that rationalize $\{q_j(\mathbf{a})\}_{j=0,\ldots,J}$ on $\mathcal{A}$. In particular, the set $\mathcal{A} \cup \{\mathbf{a}'\}$ is discrete, and therefore, it is only condition C(i) that has any empirical content for demand on $\mathcal{A} \cup \{\mathbf{a}'\}$.[6] This argument is identical to Bhattacharya, 2021 Proposition 1.

Remark 2. Allen and Rehbeck, 2019 derive bounds for $\{q_j(\mathbf{a}')\}_{j=0,\ldots,J}$ when $\mathbf{a}' \notin \mathcal{A}$ by assuming the additive structure $w_j(a_j, v_j) = w_j(a_j) + v_j$ and also assuming that $w_j(a'_j)$ is known even if $\mathbf{a}' \notin \mathcal{A}$. This is possible if $w_j(\cdot)$ and the joint distribution of unobserved heterogeneity are parametrically specified, and the values of these parameters are known from the observed choice probabilities. In contrast, the bounds in (24) do not require such parametric restrictions on the utility indices $w_j(\cdot, \cdot)$.

Change in Choice Sets: From an initial situation described by the set-up, suppose alternative J is eliminated from the choice-set. Then the choice probability $q_j(\mathbf{a} \backslash \{J\})$ of alternative $j \in \{0, 1, 2, \ldots, J-1\}$ can be obtained as follows. First the utilities $w_j(a_j, v_j)$ and the joint distribution $F_{\mathbf{v}}(v_1, ..v_{J-1}, v_J)$ are obtained by applying Lemma 1 to the original choice probabilities when the entire choice set was available. Then the joint distribution $F_{\mathbf{v}_{-J}}(v_1, ..v_{J-1})$ is obtained as

$$\lim_{v_J \nearrow \infty} F_{\mathbf{v}}(v_1, v_2, \ldots, v_{J-1}, v_J)$$

Finally, the choice probability $q_j(\mathbf{a} \backslash \{J\})$ of alternative $j \in \{0, 1, 2, \ldots, J-1\}$ is obtained as

$$\int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \prod_{k=0, k\ne j}^{J-1} 1\left\{ w_j(a_j, v_j) \ge w_k(a_k, v_k) \right\} dF_{\mathbf{v}_{-J}}(v_1, ..v_{J-1}). \tag{25}$$

If $q_0(\mathbf{a})$ does not take values in all of $[0, 1]$ on the support of the a's, then the joint distribution of the v's is identified on a subset of its support (where the CDF remains bounded away from 0 and/or 1), then one can obtain a lower bound for (25) by integrating over that smaller subset. i.e. draw v's from the identified part of the distribution, for each such draw compute $1\{w_j(a_j, v_j) \ge w_k(a_k, v_k)\}$ for all $k \ne j$, and then average over many draws.

3.4. Random coefficient models

The scope of our main result extends beyond models with additive heterogeneity, and covers the more general specification of 'random coefficients' for individual preferences which is popular in empirical IO, e.g. the mixed logit model of McFadden and Train, 2000 or BLP, 1996. In fact, McFadden and Train, 2000 show that essentially all choice probability functions generated via utility maximization can be approximated arbitrarily well by an appropriately defined mixed multinomial logit model. Such random

[6] Conditions C(ii) and C(iii) arise from the additive unobserved heterogeneity. The additivity implies rank invariance, as explained under "Further Points" following Lemma 1, above. As such, it adds no restriction on behavior in a single cross section. Further, C(ii) and C(iii) restrict *local* behavior, related to smoothness of the choice probability function, and as such, have no implication for demand prediction for an isolated, counterfactual budget set.

coefficient models posit that utility from an alternative is a linear index of its characteristics, where the coefficients in the index are individual specific. Specifically, utility of the ith individual from choosing the jth alternative is given by

$$U_{ij} = \eta_{ij0} + \sum_{k=1}^{K} z_{jk}\eta_{ik} + \eta_{ip}U\left(y_i - p_j\right),$$

where $\mathbf{z}_j = \left\{z_{j1}, ..., z_{jK}\right\}_{j=1,...,J}$ represents a vector of K observed characteristics of alternative j, and $\left(\eta_{ij0}, \eta_{i1}, ..., \eta_{iK}, \eta_{ip}\right)$ is a random coefficient vector where $\eta_{ip} > 0$ with probability 1 (reflecting non-satiation in the quantity of numeraire), and $U(\cdot)$ is a potentially nonlinear, unknown sub-utility function.[7] Then

$$\begin{aligned} & U_{ij} > U_{il} \\ & \Leftrightarrow \eta_{ij0} + \sum_{k=1}^{K} z_{jk}\eta_{ik} + \eta_{ip}U\left(y_i - p_j\right) > \eta_{il0} + \sum_{k=1}^{K} z_{lk}\eta_{ik} + \eta_{ip}U\left(y_i - p_l\right) \\ & \Leftrightarrow U\left(y_i - p_j\right) + \underbrace{\frac{\eta_{ij0}}{\eta_{ip}} + \sum_{k=1}^{K} z_{jk}\frac{\eta_{ik}}{\eta_{ip}}}_{\varepsilon_{ji}(\mathbf{z}_j)} > U\left(y_i - p_l\right) + \underbrace{\frac{\eta_{il0}}{\eta_{ip}} + \sum_{k=1}^{K} z_{lk}\frac{\eta_{ik}}{\eta_{ip}}}_{\varepsilon_{li}(\mathbf{z}_l)} \end{aligned}$$

which amounts to choice based on the utility functions $U\left(y_i - p_j\right) + \varepsilon_{ji}\left(\mathbf{z}_j\right)$. Therefore, for each realization of $\mathbf{z} = \left\{\mathbf{z}_j\right\}_{j=1,...,J}$, the conditions and thus conclusions of Theorem 1 hold; the only difference is that the choice probabilities appearing in the statement of the theorem will have to be defined conditional on $\mathbf{z} = \left\{\mathbf{z}_j\right\}_{j=1,...,J}$. Similarly, the identification argument of Sec 4.1 will work conditional on $\mathbf{z}$, implying that the joint distribution of $\left\{\varepsilon_{ji}\left(\mathbf{z}_j\right)\right\}$, $j = 1, ..., J$ is exactly identified[8] while $U(\cdot)$ may be over-identified. For example, one would expect the characteristics of alternatives viz. $\mathbf{z}$ to remain identical across consumers in a single market (e.g. the frequency of various modes of public transport are likely to be identical across individuals in the same locality). Then the $q_j(\cdot;\mathbf{z})$'s and their partial derivatives are identified via the *variation in income* y, and hence in $a_0 = y$ and $a_j \equiv y - p_j$ for $j = 1, ..., J$, across individuals in the same market and, additionally, any variation in price within and across markets with the same observed $\mathbf{z}$'s. Applying the identification argument outlined in Section 3.1, one obtains the distribution of $\left\{\varepsilon_{ji}\left(\mathbf{z}_j\right)\right\}$, $j = 1, ..., J$ conditional on each realization of $\mathbf{z}$ and the utility indices. These objects will yield bounds on choice probabilities when the budget set takes counterfactual values due to potential policy interventions, by applying (23) or (24) conditional on the $\mathbf{z}'$s.

Note further that knowledge of the distribution of the (suitably normalized) η's will allow one to bound choice probabilities when *not only the budget set but also covariates take counterfactual values*. If the number of markets is large, the distribution of random coefficients is identical in each market, and there is sufficient independent variation of the $\mathbf{z}$'s across markets, then one can identify the distribution of the normalized ηs from the distribution of the $\varepsilon_j\left(\mathbf{z}_j\right)$'s by using the Cramer-Wold theorem (cf. Billingsley 1995, Theorem 29.4, Beran and Hall, 1992). To see this, let the value of $\mathbf{z}_j$ in market m be denoted by $\mathbf{z}_j^m$, and denote $\varepsilon_{ji}\left(\mathbf{z}_j\right) = \gamma_i'\mathbf{z}_j^m$, where $\mathbf{z}_j^m$ is observed, and the object of interest is the distribution of the unobserved random coefficients γ which are the normalized values of the η's. Then, using Lemma 1, we obtain the joint distribution of $\left(\gamma'\mathbf{z}_1^m, ..., \gamma'\mathbf{z}_J^m\right)$ in market m, and therefore the marginal of $\gamma'\mathbf{z}_1^m$. Doing this in each market gives us the marginal distribution of each of the projections $\left\{\gamma'\mathbf{z}_1^m\right\}$, $m = 1, ..., M$. Now applying the approach of Beran and Hall, 1992 as $M \to \infty$ identifies the distribution of γ under appropriate regularity conditions. The precision of the corresponding estimator can be increased by using information on all J alternatives, i.e. $\left\{\gamma'\mathbf{z}_j^m\right\}$, $m = 1, ..., M$, $j = 1, 2, ..., J$.

We conclude this subsection with the observation that Lemma 1 also applies to more general models e.g. where utilities are given by

$$U_{ij} = \eta_{ip}U\left(y_i - p_j, \mathbf{z}_j\right) + \varepsilon_{ji}\left(\mathbf{z}_j\right), \tag{26}$$

where $\eta_{ip} > 0$ with probability 1, and the unobserved $\varepsilon_{ji}\left(\mathbf{z}_j\right)$ is not necessarily linear in $\mathbf{z}_j$. Condition (ii) of Theorem 1, conditional on observed covariates, is therefore a testable implication of all such models.

4. Conclusion

This paper provides a unified analysis of integrability and identification in multinomial discrete choice models. It establishes closed-form shape-restrictions on choice-probability functions, under which multinomial choice probabilities can be rationalized via random utility models. These conditions are shown to be necessary and sufficient for the additive random utility model of McFadden. Our results apply equally to random coefficient models like mixed logit – widely used in IO applications – because conditional on observed characteristics, these are observationally equivalent to models with additive heterogeneity. Our theoretical results are obtained via application of the classical theory of partial differential equations, whose use in economics and econometrics is relatively

[7] If $U(\cdot)$ is linear, then income drops out of choice probabilities, which is a strong and testable restriction.

[8] Fox, 2021 derives identification results for distribution of random coefficients in multinomial choice models, in presence of a covariate with large support. Our condition C(iii) can be thought of as an alternative to the large support assumption which also delivers identification.

novel. The key empirical implications of our results are that they lead to (a) nonparametric identification of random utility models using economic theory as opposed to statistical assumptions, (b) specification of multinomial choice models in applied work that is consistent with economic theory while allowing for nonparametric utility functions, unobserved heterogeneity and income-effects, and (c) calculation of theory-consistent nonparametric bounds for demand and welfare on counterfactual budget sets, e.g. those arising from price change (say, due to a tax or subsidy, firm-mergers etc.) and changes in the number of available alternatives.

5. Appendix

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

Result 1 (cf. Zachmanglou and Thoe 1986 Theorem 4.1): Consider the linear homogeneous PDE

$$\sum_{j=1}^{J} g_j(\mathbf{x}) \frac{\partial \sigma(\mathbf{x})}{\partial x_j} = 0, \tag{27}$$

where $\mathbf{x} \in \mathbb{R}^J$, $g_j(\cdot)$ are C^1 functions from $\mathbb{R}^J$ to $\mathbb{R}$, and do not vanish simultaneously on a domain Ω, with $g_1(\mathbf{x}) \equiv 1$. Then a general solution to (27) is given by

$$\sigma(\mathbf{x}) = \phi\left(h_2(\mathbf{x}), \ldots h_J(\mathbf{x})\right), \tag{28}$$

where $\phi(\cdot)$ is *any* arbitrary C^1 function, and for $j = 2, \ldots J$, the functions $h_j(\cdot)$ satisfy that $h_j(\mathbf{x}) = c_j$, where c_js are constants are general solutions to the ordinary differential equations

$$\frac{dx_1}{1} = \frac{dx_2}{g_2(\mathbf{x})} = \ldots = \frac{dx_J}{g_J(\mathbf{x})}. \tag{29}$$

Discussions: The ODE (29) are known as the "characteristic equations" of the linear PDE (27), and existence of a solution to the PDE (27) amounts to existence of a solution of the ODE (29), cf. Courant, 1962, Chapter I.5, II.2 and Zachmanglou and Thoe 1986 Theorem 4.1. In (28), $\phi(\cdot)$ can be chosen to be strictly increasing in both arguments. A unique choice of $\phi(\cdot)$ is pinned down by boundary conditions; in our application, these amount to equating $\phi\left(h_2(\mathbf{x}), \ldots h_J(\mathbf{x})\right)$ to observed choice probability functions. For a geometric idea behind this method of solving PDEs, we refer the reader to Courant and Hilbert 1962, page 30.

The second result states the Picard-Lindelöf theorem that establishes conditions for existence of a unique solution to a first-order ODE such as the characteristic ODEs (29). See, for example, *Coddington, 1961, Chapter 5.8, Corollary 1.*

Result 2 (Picard-Lindelöf Theorem): *Suppose that a function $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is continuous and Lipschitz on $\mathbb{R}$. Then the ordinary differential equation $n'(x) = g(x, n(x))$ with initial condition $n(x_0) = n_0$ has a unique solution $n(\cdot) : \mathbb{R} \to \mathbb{R}$* on any interval I containing x_0, *with $n(\cdot)$ being C^1.*

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

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

Example: As a closed-form example of the above method, consider the PDE

$$\frac{\partial \sigma(x_1, x_2, x_3)}{\partial x_1} + x_2 \frac{\partial \sigma(x_1, x_2, x_3)}{\partial x_2} + x_3 \frac{\partial \sigma(x_1, x_2, x_3)}{\partial x_3} = 0.$$

The characteristic curve (29) will satisfy the ODEs

$$\frac{dx_1}{1} = \frac{dx_2}{x_2} = \frac{dx_3}{x_3},$$

$$\text{or, equivalently, } \frac{dx_2}{dx_1} = x_2,\ \frac{dx_3}{dx_1} = x_3,$$

implying general solutions of the form

$$\sigma(x_1, x_2, x_3) = \phi\left(\underbrace{x_1 - \ln(x_2)}_{h_2(\mathbf{x})},\ \underbrace{x_1 - \ln(x_3)}_{h_3(\mathbf{x})}\right),$$

for any arbitrary smooth $\phi(\cdot, \cdot)$.

Proof of Lemma 1

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

$$t_{j0}\left(a_j, a_0\right) \equiv \frac{\partial}{\partial a_0} q_j(\mathbf{a}) / \frac{\partial}{\partial a_j} q_0(\mathbf{a}) > 0, \tag{30}$$

where the positivity is a consequence of condition C(i).

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

$$\frac{\partial}{\partial a_0} q_0(\mathbf{a}) + \sum_{j=1}^{J} \frac{\partial}{\partial a_0} q_j(\mathbf{a}) = 0. \tag{31}$$

Substituting (30) in (31), we get the linear, homogeneous, partial differential equation in $q_0(\cdot)$:

$$\frac{\partial}{\partial a_0} q_0(\mathbf{a}) + \sum_{j=1}^{J} \frac{\partial}{\partial a_j} q_0(\mathbf{a}) \times t_{j0}\left(a_j, a_0\right) = 0, \tag{32}$$

which is exactly of the form (27), with $\mathbf{x} = \left(a_0, a_1, ..., a_J\right)$, $g_j(\mathbf{x}) = t_{j0}\left(a_j, a_0\right)$ $j = 0, ...J$ and $q_0(\mathbf{a}) = \sigma(\mathbf{x})$.

Therefore, this PDE can be solved via the method of characteristics (see Result 1 above), giving the characteristic Ordinary Differential Equations:

$$\frac{d a_j}{d a_0} = t_{j0}\left(a_j, a_0\right), \tag{33}$$

for $j = 1, ..., J$. Using the Picard-Lindelöf theorem (Result 2 above), we obtain the general solutions of the ODE (33) given by $\omega_j\left(a_j, a_0\right)$ =constant,[9] where $\omega_j\left(a_j, a_0\right)$ is differentiable, and satisfies

$$\frac{\partial \omega_j\left(a_j, a_0\right)}{\partial a_0} + \frac{\partial \omega_j\left(a_j, a_0\right)}{\partial a_j} t_{j0}\left(a_j, a_0\right) = 0. \tag{34}$$

Recall that C(i) implies $t_{j0}\left(a_j, a_0\right) > 0$, so from (34), we have that $\frac{\partial \omega_j\left(a_j, a_0\right)}{\partial a_0} / \frac{\partial \omega_j\left(a_j, a_0\right)}{\partial a_j} < 0$ and therefore, we can choose a solution $\omega_j\left(a_j, a_0\right)$ that is strictly increasing in a_0 and strictly decreasing in a_j while satisfying (34).

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

$$q_0(\mathbf{a}) = H_0\left(\omega_1\left(a_1, a_0\right), \omega_2\left(a_2, a_0\right), ..., \omega_J\left(a_J, a_0\right)\right), \tag{35}$$

where $H_0(\cdot)$ can be chosen to be strictly increasing and C^1 in each argument, with continuous Jth order cross partial derivatives and taking values in $[0, 1]$. Since $q_0(\mathbf{a})$ is observed, the exact functional form of $H_0(\cdot)$ is pinned down by (35), for any set of solutions $\omega_j(\cdot, \cdot)$ to the ODEs (33). This corresponds to the so-called "initial condition" in the PDE nomenclature. In particular, given any a_0, the value of $H_0\left(x_1, x_2, ...x_J\right)$ at any vector $\left(x_1, x_2, ...x_J\right)$ is given by

$$H_0\left(x_1, x_2, ...x_J\right) = q_0\left(a_0, b_1\left(x_1, a_0\right), ...b_J\left(x_J, a_0\right)\right), \tag{36}$$

where $b_j\left(x_j, a_0\right)$ is defined by the solution b to

$$\omega_j\left(b, a_0\right) = x_j \tag{37}$$

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

$$\begin{aligned}
& q_0\left(a_0, b_1\left(x_1, a_0\right), ...b_J\left(x_J, a_0\right)\right) \\
& = H_0\left(\omega_1\left(b_1\left(x_1, a_0\right), a_0\right), \omega_2\left(b_2\left(x_2, a_0\right), a_0\right), ..., \omega_J\left(b_J\left(x_J, a_0\right), a_0\right)\right) \text{ from (35)} \\
& = H_0\left(x_1, x_2, ...x_J\right) \\
& \overset{\text{from (36)}}{=} H_0\left(\omega_1\left(b_1\left(x_1, a_0'\right), a_0'\right), \omega_2\left(b_2\left(x_2, a_0'\right), a_0'\right), ..., \omega_J\left(b_J\left(x_J, a_0'\right), a_0'\right)\right) \\
& \overset{\text{from (35)}}{=} q_0\left(a_0', b_1\left(x_1, a_0'\right), ...b_J\left(x_J, a_0'\right)\right).
\end{aligned} \tag{38}$$

Having obtained the $\omega_j(\cdot, \cdot)$'s from (33) and (35), let $\bar{\mathcal{V}}_j$ denote the co-domain of $\omega_j(\cdot, \cdot)$ (i.e. the set of images of $\omega_j(\cdot, \cdot)$ as the arguments of $\omega_j(\cdot, \cdot)$ vary over their support) and for each $j = 1, ...J$ and $v \in \bar{\mathcal{V}}_j$, construct the function $w_j\left(a_j, v_j\right)$ by inversion, i.e.

$$w_j\left(a_j, v_j\right) = \left\{a_0 : \omega_j\left(a_j, a_0\right) = v_j\right\}, \tag{39}$$

[9] As a clarifying example, the ODE $\frac{dy}{dx} = 2x$ has the solution $y = x^2 + const$, or equivalently, $\omega(x, y) \equiv y - x^2$ =constant.

which is well-defined by the inverse function theorem as $\omega_j(a_j, a_0)$ is invertible in its second argument. Note that by construction, $w_j(a_j, v_j)$ is strictly increasing and continuous in a_j. 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

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

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

$$F(v_1, ..., v_J) = q_0(a_0, a_1, ..., a_J), \tag{40}$$

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 (35) and (38) that this function is well-defined. The above CDF implies the density function $f : \mathcal{V} \to \mathbb{R}^+$:

$$\begin{aligned}
& f(v_1, ..., v_J) \\
&= \frac{\frac{\partial^J}{\partial a_1 ... \partial a_J} q_0(a_0, a_1, ..., a_J) |_{v_j=\omega_j(a_j,a_0),\ j=1,...J}}{\prod_{j=1}^{J} \frac{\partial}{\partial a_j}\omega_j(a_j, a_0) |_{v_j=\omega_j(a_j,a_0),\ j=1,...J}} \qquad (41) \\
&= \frac{\frac{\partial^{J-1}}{\partial a_1 ... \partial a_{k-1}\partial a_{k+1} ... \partial a_J} \frac{\partial}{\partial a_k} q_0(a_0, a_1, ..., a_J) |_{v_j=\omega_j(a_j,a_0),\ j=1,...J}}{\prod_{j=1}^{J} \frac{\partial}{\partial a_j}\omega_j(a_j, a_0) |_{v_j=\omega_j(a_j,a_0),\ j=1,...J}}, \text{ for any } k \in \{1, ..., J\} \\
&= \frac{\frac{\partial^{J-1}}{\partial a_1 ... \partial a_{k-1}\partial a_{k+1} ... \partial a_J}\left[-\underbrace{\frac{\frac{\partial \omega_k(a_k,a_0)}{\partial a_k}}{\frac{\partial \omega_k(a_k,a_0)}{\partial a_0}}}_{\text{does not depend on } a_1...a_{k-1},a_{k+1}...a_J} \times \frac{\partial}{\partial a_0} q_k(a_0, a_1, ..., a_J) \right] \Bigg|_{v_j=\omega_j(a_j,a_0)}}{\prod_{j=1}^{J} \frac{\partial}{\partial a_j}\omega_j(a_j, a_0) \Big|_{v_j=\omega_j(a_j,a_0),\ j=1,...J}}, \text{ from (19)} \\
&= -\frac{\frac{\partial^{J-1}}{\partial a_1 ... \partial a_{k-1}\partial a_{k+1} ... \partial a_J} \frac{\partial}{\partial a_0} q_k(a_0, a_1, ..., a_J) |_{v_j=\omega_j(a_j,a_0),\ j=1,...J}}{\frac{\frac{\partial}{\partial a_0}\omega_k(a_k,a_0)}{\frac{\partial}{\partial a_k}\omega_k(a_k,a_0)} \times \prod_{j=1}^{J} \frac{\partial}{\partial a_j}\omega_j(a_j, a_0) |_{v_j=\omega_j(a_j,a_0),\ j=1,...J}} \\
&= -\frac{\frac{\partial^{J-1}}{\partial a_1 ... \partial a_{k-1}\partial a_{k+1} ... \partial a_J} \frac{\partial}{\partial a_0} q_k(a_0, a_1, ..., a_J) |_{v_j=\omega_j(a_j,a_0),\ j=1,...J}}{\frac{\partial}{\partial a_0}\omega_k(a_k, a_0) \times \prod_{j=1, j\neq k}^{J} \frac{\partial}{\partial a_j}\omega_j(a_j, a_0) |_{v_j=\omega_j(a_j,a_0),\ j=1,...J}}. \qquad (42)
\end{aligned}$$

Condition C(iii) implies $(-1)^J \times d^J q_k(a_0, ..., a_J)/da_0...da_J$ is non-negative 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 $\mathcal{V}$, each of the above expressions has numerator and denominator of the same sign, and is thus non-negative.

We now show that the above construction of $w_j(\cdot,\cdot)$ (cf. (39)) and the joint density of heterogeneity (38) and (42) will indeed produce the original choice probabilities. To see this for alternative 1, consider the integral

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

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

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

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

$$=(-1)^{J-1}\times\int_{a_0}^{\infty}\int_{a_2}^{\infty}...\int_{a_J}^{\infty}\left\{-\frac{\partial^J}{\partial x_1\partial x_2...\partial x_J}q_1\left(x_1,a_1,x_2,...x_J\right)\right\}dx_J...dx_2dx_1\text{, by (42)}$$

$$=(-1)^{J}\times\int_{a_0}^{\infty}\int_{a_2}^{\infty}...\int_{a_J}^{\infty}\left\{\frac{\partial^J}{\partial x_1\partial x_2...\partial x_J}q_1\left(x_1,a_1,x_2,...x_J\right)\right\}dx_J...dx_2dx_1$$

$$=\int_{\infty}^{a_0}\int_{\infty}^{a_2}...\int_{\infty}^{a_J}\left\{\frac{\partial^J}{\partial x_1\partial x_2...\partial x_J}q_1\left(x_1,a_1,x_2,...x_J\right)\right\}dx_J...dx_2dx_1$$

$$=q_1\left(a_0,a_1,a_2,...a_J\right). \tag{43}$$

Exactly analogous steps for $j=2,...J$, and using (42), lead to the conclusion that for all $j\geq 1$,

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

Also, note that

$$\int 1\left\{a_0\geq\max_{k\in\{1,2,...J\}}w_k\left(a_k,v_k\right)\right\}f\left(v_1,v_2,...,v_J\right)dv_1...dv_J$$
$$=\int_0^{\omega_1(a_1,a_0)}...\int_0^{\omega_J(a_J,a_0)}f\left(v_1,v_2,...,v_J\right)dv_J...dv_1$$

substitute $v_j\to x_j$ satisfying $v_j=\omega_j\left(x_j,a_0\right)$

$$=\int_{a_1}^{\infty}\int_{a_2}^{\infty}...\int_{a_J}^{\infty}f\left(\omega_1\left(x_1,a_0\right),...,\omega_J\left(x_J,a_0\right)\right)\left|\frac{\partial\omega_1\left(x_1,a_0\right)}{\partial x_1}...\frac{\partial\omega_J\left(x_J,a_0\right)}{\partial x_J}\right|dx_J...dx_1$$

$$=\int_{a_1}^{\infty}\int_{a_2}^{\infty}...\int_{a_J}^{\infty}(-1)^J\times f\left(\omega_1\left(x_1,a_0\right),...,\omega_J\left(x_J,a_0\right)\right)\frac{\partial\omega_1\left(x_1,a_0\right)}{\partial x_1}...\frac{\partial\omega_J\left(x_J,a_0\right)}{\partial x_J}dx_J...dx_1$$

$$=\int_{\infty}^{a_1}...\int_{\infty}^{a_J}\frac{\partial^J}{\partial\alpha_1...\partial\alpha_J}q_0\left(a_0,\alpha_1,...\alpha_J\right)|_{\alpha_1=x_1,...\alpha_J=x_J}dx_J...dx_1\text{, by (41)}$$

$$=q_0\left(a_0,a_1,...a_J\right).$$

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 (38) would indeed produce the choice probabilities $\left\{q_j\left(\cdot,...\cdot\right)\right\}$ for each $j=0,1,...J$. □

Proof of Theorem 1

Proof. Necessity: Observe that the choice probability for the 0th alternative is given by

$$q_0\left(\mathbf{a}\right)$$
$$=\Pr\left(\cap_{j\neq 0}\left\{h_0\left(a_0\right)+\varepsilon_0>h_j\left(a_j\right)+\varepsilon_j\right\}\right)$$
$$=\Pr\left[\cap_{j\neq 0}\left\{h_0\left(a_0\right)-h_j\left(a_j\right)>\varepsilon_j-\varepsilon_0\right\}\right]$$
$$=\int_{-\infty}^{\infty}\int_{-\infty}^{(h_0(a_0)-h_1(a_1))+\varepsilon_0}...\int_{-\infty}^{(h_0(a_0)-h_J(a_J))+\varepsilon_0}g\left(\varepsilon\right)\prod_{j=0}^{J}d\varepsilon_j. \tag{44}$$

Therefore, by the first fundamental theorem of calculus and chain-rule,

$$\frac{\partial}{\partial a_1} q_0(\mathbf{a}) = -h_1'(a_1) \int_{-\infty}^{\infty} \int_{-\infty}^{\substack{\varepsilon_0 \\ +h_0(a_0) \\ -h_2(a_2)}} \dots \int_{-\infty}^{\substack{\varepsilon_0 \\ +h_0(a_0) \\ -h_J(a_J)}} g\left(\varepsilon_0, h_0(a_0) - h_1(a_1) + \varepsilon_0, \varepsilon_2, \dots \varepsilon_J\right) \prod_{\substack{j=0 \\ j \neq 1}}^{J} d\varepsilon_j. \tag{45}$$

Similarly,

$$q_1(\mathbf{a}) = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\substack{\varepsilon_1 \\ +h_1(a_1) \\ -h_0(a_0)}} \dots \int_{-\infty}^{\substack{\varepsilon_1 \\ +h_1(a_1) \\ -h_J(a_J)}} g(\varepsilon) \prod_{\substack{j=0 \\ j \neq 1}}^{J} d\varepsilon_j \right) d\varepsilon_1,$$

implying by the first fundamental theorem and chain-rule that

$$\begin{aligned}
&\frac{\partial}{\partial a_0} q_1(\mathbf{a}) \\
&= -h_0'(a_0) \int_{-\infty}^{\infty} \int_{-\infty}^{\substack{h_1(a_1) \\ -h_2(a_2)+\varepsilon_1}} \dots \int_{-\infty}^{\substack{h_1(a_1) \\ -h_J(a_J)+\varepsilon_1}} g\begin{pmatrix} h_1(a_1) - h_0(a_0) + \varepsilon_1, \\ \varepsilon_1, \varepsilon_2, \dots \varepsilon_J \end{pmatrix} d\varepsilon_J \dots d\varepsilon_2 d\varepsilon_1 \qquad (46) \\
&\overset{(1)}{=} -h_0'(a_0) \int_{-\infty}^{\infty} \int_{-\infty}^{\substack{s_0+h_0(a_0) \\ -h_2(a_2)}} \dots \int_{-\infty}^{s_0+h_0(a_0)-h_J(a_J)} g\left(s_0, s_0 - h_1(a_1) + h_0(a_0), \varepsilon_2, \dots \varepsilon_J\right) d\varepsilon_J \dots d\varepsilon_2 ds_0 \\
&= \frac{h_0'(a_0)}{h_1'(a_1)} \frac{\partial}{\partial a_1} q_0(\mathbf{a}), \text{ using (45)},
\end{aligned}$$

where the second equality $\overset{(1)}{=}$ follows by substituting $s_0 = h_1(a_1) - h_0(a_0) + \varepsilon_1$ in (46). The same argument can be repeated for any other pair of alternatives $l \neq k$, to obtain

$$\frac{\frac{\partial}{\partial a_k} q_l(\mathbf{a})}{\frac{\partial}{\partial a_l} q_k(\mathbf{a})} = \frac{h_k'(a_k)}{h_l'(a_l)}, \tag{47}$$

for all $\mathbf{a}$, and it is clear that the RHS of (47) has the same form as stated in condition C(ii). Next, Condition C(i) follows from the fact that

$$\begin{aligned}
q_j(a_0, a_1, \dots, a_J) &= \Pr\left[\cap_{k \neq j} 1\left\{h_j(a_j) + v_j \geq h_k(a_k) + v_k\right\}\right] \\
&= \Pr\left[\cap_{k \neq j} 1\left\{v_j - v_k \geq h_k(a_k) - h_j(a_j)\right\}\right]
\end{aligned}$$

where the functions $h_j(\cdot)$ are strictly increasing, and $v_j - v_k$ have a continuous joint density. Finally, C(iii) holds because

$$\begin{aligned}
&\frac{\partial^J}{\partial a_0 \partial a_1 \dots \partial a_{r-1} \partial a_{r+1} \dots \partial a_J} q_r(\mathbf{a}) \\
&= \frac{\partial^J}{\partial a_0 \partial a_1 \dots \partial a_{r-1} \partial a_{r+1} \dots \partial a_J} \left(\Pr\left[\cap_{k \neq r} 1\left\{v_r - h_k(a_k) + h_r(a_r) \geq v_k\right\}\right]\right) \\
&= \frac{\partial^J}{\partial a_0 \partial a_1 \dots \partial a_{r-1} \partial a_{r+1} \dots \partial a_J} \left(\int_{-\infty}^{\infty} \int_{-\infty}^{v_r - h_0(a_0) + h_r(a_r)} \dots \int_{-\infty}^{v_r - h_J(a_J) + h_r(a_r)} f(v_0, v_1, \dots, v_J) \right) dv_J \dots dv_0 dv_r \\
&= f\left(v_r - h_0(a_0) + h_r(a_r), \dots, v_r, \dots, v_r - h_J(a_J) + h_r(a_r)\right) \times (-1)^J \times \prod_{k \neq r} h_k'(a_k),
\end{aligned}$$

where the first term denotes the joint density of $\left(v_0, \ldots v_J\right)$ evaluated at the arguments specified, and the last product term is non-negative because the $h_j(\cdot)$'s are strictly increasing and continuously differentiable.

Sufficiency: WLOG take $m=0$, and let $\bar{G}_j\left(a_j\right)$ and $\bar{G}_0\left(a_0\right)$ be the primitive integrals of $G_j\left(a_j\right)$ and $G_0\left(a_0\right)$, i.e. $\frac{d}{da_j}\bar{G}_j\left(a_j\right)=G_j\left(a_j\right)$ and $\frac{d}{da_0}\bar{G}_0\left(a_0\right)=G_0\left(a_0\right)$; note that $\bar{G}_j\left(a_j\right)$ and $\bar{G}_0\left(a_0\right)$ are strictly increasing and continuous since they have strictly positive derivatives. Then, by exactly analogous steps that led to (35) in the proof of Lemma 1, we have that condition (ii) of Theorem 1, viz. $\frac{\partial}{\partial a_m}q_j(\mathbf{a})/\frac{\partial}{\partial a_j}q_m(\mathbf{a})=G_m\left(a_m\right)/G_j\left(a_j\right)$ has a general solution of the form

$$q_0(\mathbf{a})=H\left(\bar{G}_1\left(a_1\right)-\bar{G}_0\left(a_0\right), \ldots \bar{G}_J\left(a_J\right)-\bar{G}_0\left(a_0\right)\right),$$

where $H(\cdot)$ is an arbitrary smooth function mapping $\mathbb{R}^J \rightarrow [0,1]$. In particular, we can take $H(\cdot)$ to be nondecreasing in each argument, and we have that $\bar{G}_j(\cdot)$, $j=0,\ldots,J$ are strictly increasing and continuous. Following exactly analogous steps to the proof of Lemma 1, we get that $q_j(\mathbf{a})$ is rationalized by the utility functions $w_0\left(a_0,\eta\right)=a_0$, $w_j\left(a_j,\eta\right)=\bar{G}_0^{-1}\left(\bar{G}_j\left(a_j\right)-v_j\right)$, with the CDF for the joint distribution of the unobserved heterogeneity $\eta \equiv \left(v_1,\ldots,v_J\right)$ given by

$$F_\eta\left(v_1,v_2,\ldots,v_J\right)=q_0\left(a_0,\bar{G}_1^{-1}\left(\bar{G}_0\left(a_0\right)+v_1\right),\ldots\bar{G}_J^{-1}\left(\bar{G}_0\left(a_0\right)+v_J\right)\right) \tag{48}$$

$$=H\left(v_1,\ldots,v_J\right) \tag{49}$$

Just as in (38), the choice of a_0 is immaterial here. Note further that the above model is observationally equivalent to one where utilities are given by $W_0\left(a_0,\eta\right)=\bar{G}_0\left(a_0\right)$, $W_j\left(a_j,\eta\right)=\bar{G}_j\left(a_j\right)-v_j$, $j=1,\ldots J$, with the joint CDF of $\eta\equiv\left(v_1,\ldots,v_J\right)$ still given by (49). This is precisely the ARUM model. That these distributions imply

$$q_j(\mathbf{a})=\Pr[\cap_{k\neq j}W_j\left(a_j,\eta\right)\geq W_k\left(a_k,\eta\right)]$$

for all $j=0,\ldots,J$ can be established following the exact same steps as in the proof of Lemma 1 above, with $\omega_j\left(a_j,a_0\right)$ replaced by $\bar{G}_j\left(a_j\right)-\bar{G}_0\left(a_0\right)$ everywhere. □

CRediT authorship contribution statement

Debopam Bhattacharya: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

None

Data availability

No data was used for the research described in the article.

References

Allen, R., Rehbeck, J., 2019. Identification with additively separable heterogeneity. Econometrica 87 (3), 1021–1054.
Armstrong, M., Vickers, J., 2015. Which demand systems can be generated by discrete choice? J. Econ. Theory 158, 293–307.
Beran, R., Hall, P., 1992. Estimating coefficient distributions in random coefficient regressions. Ann. Stat. 20 (4), 1970–1984.
Berry, S.T., Haile, P.A., 2021. Foundations of Demand Estimation (No. w29305). National Bureau of Economic Research.
Bhattacharya, D., 2018. Empirical welfare analysis for discrete choice: some general results. Quant. Econ. 9 (2), 571–615.
Bhattacharya, D., 2021. The empirical content of binary choice models. Econometrica 89 (1), 457–474.
Block, H.D., Marschak, J., 1960. Random orderings and stochastic theories of response. In: Olkin, I., et al. (Eds.), Contributions to Probability and Statistics. Stanford University Press, Stanford.
Blundell, R., Powell, James L., 2003. Endogeneity in nonparametric and semiparametric regression models. Econom. Soc. Monogr. 36, 312–357.
Blundell, R.W., Powell, J.L., 2004. Endogeneity in semiparametric binary response models. Rev. Econ. Stud. 71 (3), 655–679.
Chernozhukov, V., Fernández-Val, I., Newey, W.K., 2019. Nonseparable multinomial choice models in cross-section and panel data. J. Econom. 211 (1), 104–116.
Coddington, E.A., 1961. An Introduction to Ordinary Differential Equations. Dover Publishing, New York.
Courant, R., 1962. Methods of Mathematical Physics, vol. 2. Interscience, New York.
Daly, A., Zachary, S., 1978. Improved multiple choice models. In: Hensher, D., Dalvi, Q. (Eds.), Identifying and Measuring the Determinants of Mode Choice. Teakfields, London.
Deb, R., Kitamura, Y., Quah, J.K.H., Stoye, J., 2021. Revealed price preference: theory and empirical analysis. preprint arXiv:1801.02702.
Dette, H., Hoderlein, S., Neumeyer, N., 2016. Testing multivariate economic restrictions using quantiles: the example of Slutsky negative semidefiniteness. J. Econom. 191 (1), 129–144.
Falmagne, J.C., 1978. A representation theorem for finite random scale systems. J. Math. Psychol. 18 (1), 52–72.
Fosgerau, M., McFadden, D., Bierlaire, M., 2013. Choice probability generating functions. J. Choice Model. 8, 1–18.
Fox, J.T., 2021. A note on nonparametric identification of distributions of random coefficients in multinomial choice models. Ann. Econ. Statist. 142, 305–310.
Gandhi, A., Nevo, A., 2021. Empirical models of demand and supply in differentiated products industries. Handbook of industrial organization, vol. 4, issue 1. Elsevier, pp. 63–139.
Hausman, J.A., Newey, W.K., 2016. Individual heterogeneity and average welfare. Econometrica 84 (3), 1225–1248.

Imbens, G.W., Newey, W.K., 2009. Identification and estimation of triangular simultaneous equations models without additivity. Econometrica 77 (5), 1481–1512.
Kitamura, Y., Stoye, J., 2016. Nonparametric analysis of random utility models. Econometrica 86 (6), 1883–1909.
Koning, R.H., Ridder, G., 2003. Discrete choice and stochastic utility maximization. Econom. J. 6 (1), 1–27.
Lewbel, A., 2000. Semiparametric qualitative response model estimation with unknown heteroskedasticity or instrumental variables. J. Econom. 97 (1), 145–177.
Lewbel, A., 2001. Demand systems with and without errors. Am. Econ. Rev. 91 (3), 611–618.
Manski, C.F., 2007. Partial identification of counterfactual choice probabilities. Int. Econ. Rev. 48 (4), 1393–1410.
Mas-Colell, A., Whinston, M.D., Green, J.R., 1995. Microeconomic Theory. Oxford University Press, New York.
Matzkin, R., 1993. Nonparametric identification and estimation of polychotomous choice models. J. Econom. 58 (1–2), 137–168.
Matzkin, R.L., 2007. Nonparametric identification. Handb. Econom. 6, 5307–5368.
McFadden, D., 1978. Modeling the choice of residential location. Transp. Res. Rec. 673.
McFadden, D., 1981. Econometric models of probabilistic choice. In: Structural Analysis of Discrete Data with Econometric Applications, 198272.
McFadden, D., Richter, M.K., 1990. Stochastic rationality and revealed stochastic preference. In: Preferences, Uncertainty, and Optimality, Essays in Honor of Leo Hurwicz. Westview Press, pp. 161–186.
McFadden, D., Train, K., 2000. Mixed MNL models for discrete response. J. Appl. Econom. 15 (5), 447–470.
Molinari, F., 2020. Microeconometrics with partial identification. Handb. Econom. 7 (Part A), 355–486.
Tebaldi, P., Torgovitsky, A., Yang, H., 2023. Nonparametric estimates of demand in the California health insurance exchange. Econometrica 91 (1), 107–146.
Varian, H., 1992. Microeconomic Analysis, 3rd ed. W. W. Norton, New York, pp. 120–121.