A note on identification in discrete choice models with partial observability

Fosgerau, Mogens and Ranjan, Abhishek

Danish Technical University

2017

A note on identification in discrete choice models with partial observability

Mogens Fosgerau Abhishek Ranjan*

November 24, 2016

Abstract

This note establishes a new identification result for additive random utility discrete choice models (ARUM). A decision-maker associates a random utility $U_j + m_j$ to each alternative in a finite set $j \in \{1, ..., J\}$, where $\mathbf{U} = \{U_1, ..., U_J\}$ is unobserved by the researcher and random with an unknown joint distribution, while the perturbation $\mathbf{m} = (m_1, ..., m_J)$ is observed. The decision-maker chooses the alternative that yields the maximum random utility, which leads to a choice probability system $\mathbf{m} \to (\Pr(1|\mathbf{m}), ..., \Pr(J|\mathbf{m}))$. Previous research has shown that the choice probability system is identified from the observation of the relationship $\mathbf{m} \to \Pr(1|\mathbf{m})$. We show that the complete choice probability system is identified from observation of a relationship $\mathbf{m} \to \sum_{j=1}^{s} \Pr(j|\mathbf{m})$, for any $s < J$. That is, it is sufficient to observe the aggregate probability of a group of alternatives as it depends on $\mathbf{m}$. This is relevant for applications where choices are observed aggregated into groups while prices and attributes vary at the level of individual alternatives.

1 Introduction

This note establishes a new identification result for additive random utility discrete choice models, showing that the complete system of choice probabilities is identified from observation of the joint probability for a subset of the alternatives as a function of a vector of location shifts.

A random utility model (RUM) associates a vector $\mathbf{U} = (U_1, ..., U_J)$ of random utilities with a choice set consisting of J alternatives. A decision-maker receives a realization of the random utility vector and chooses the alternative with the maximum utility. If the joint distribution of utility is absolutely continuous, then this induces a unique multinomial choice probability

*Technical University of Denmark, Denmark. Fosgerau:mfos@dtu.dk, Ranjan: abhra@dtu.dk

vector, which can be computed by the analyst given knowledge of the joint distribution of utility. An additive random utility model (ARUM) is a random utility model where the random utility vector is perturbed by a deterministic vector $\mathbf{m}$ such that the decision-maker maximizes $\mathbf{m} + \mathbf{U}$ and the choice probability vector becomes a function of $\mathbf{m}$, i.e. the choice probability system

$$\mathbf{m} \to (\Pr(1|\mathbf{m}), ..., \Pr(J|\mathbf{m})). \tag{1.1}$$

In applications, the perturbed random utility vector is parametrized to depend on observable variables.

Matzkin (2007) showed that an ARUM is identified from the probability of a single alternative. That means that it is possible to determine the whole choice probability system (1.1) from the observation of a function $\mathbf{m} \to \Pr(j|\mathbf{m})$ that relates the probability of a single alternative j to the perturbation vector $\mathbf{m}$.

In this paper, we extend this result to the case where one observes the probability that the choice is in a set of alternatives that is any proper subset of the choice set. An example may be where the researcher observes prices and characteristics for all the different car models on a market, but where he/she only observes demand , e.g., at the level of brands. In such situations it is important to know what is identified from the data and what identification relies on parametric model specification.

The result in this paper is a nonparametric identification result. Such results are useful to establish, on the one hand, the limits of what can be learnt from data and, on the other hand, to develop nonparametric estimators.

Random utility models, and most often additive random utility models, have been extensively used in economics and other social science fields since the pioneering work of McFadden (1974). Amemiya (1981) and Maddala (1983) discuss an extensive list of applications of this model. These include the choice of mode of transportation, choice of occupation, and choice of residence.

Section 2 first presents some preliminaries. Section 3 establishes identification from observation of the probability of a single alternative. This is expanded in Section 4 to the case where the probability of a set of alternatives is observed. Section 5 establishes precisely that the choice probability vector only identifies the distribution of random utility up to a univariate random variable added to all components of the random utility vector. Section 6 shows a way to construct ARUM with a nested structure corresponding to a partitioning of choice alternatives into groups. This construction relies on the function that relates expected maximum utility to the perturbation $\mathbf{m}$. Section 7 concludes. For the exposition of the theory of ARUM we have drawn on an unpublished lecture note written by Dan McFadden (McFadden, 2014).

2 Preliminaries

A decision maker faces a finite choice set of alternatives denoted by $C = \{1, \cdots, J\}$. Let S be any non-empty proper subset of the choice set C, $\emptyset \neq S \subsetneq C$. Without loss of generality, we let S comprise the first s elements of C, where $1 \leq s < J$.

Vectors are written in bold. The vector $\mathbf{m} = (m_1, \cdots, m_J) \in \mathbb{R}^J$ has a component for each alternative in C. Let $\mathbf{m}_{-s} = (m_{s+1}, \cdots, m_J)$ denote the vector $\mathbf{m}$ excluding the components of S. Then $\mathbf{m}_{-s} - k = (m_{s+1} - k, \cdots, m_J - k)$ denotes a vector with the scalar k subtracted from every component of $\mathbf{m}_{-s}$. The notation $(\mathbf{m}_s, \mathbf{m}_{-s})$ will be used for the vector $\mathbf{m}$ when it is convenient.

Let $\sigma = (\sigma_1, \cdots, \sigma_J)$ denote a permutation of $(1, \cdots, J)$, and let $\sigma_{:k} = (\sigma_1, \cdots, \sigma_k)$ denote the first k elements of σ. Let

$$\nabla_{\sigma:k} F(\mathbf{u}) \equiv \frac{\partial^k F(u_1, \cdots, u_J)}{\partial u_{\sigma_1} \cdots \partial u_{\sigma_k}}$$

denote the mixed partial derivative of a real-valued function F on $\mathbb{R}^J$ with respect to the variables in $\sigma_{:k}$. Similarly, denote

$$\nabla_{12\cdots J} F(\mathbf{u}) \equiv \frac{\partial^J F(u_1, \cdots, u_J)}{\partial u_1 \cdots \partial u_J}.$$

Definition 1 *A (complete) choice probability system (CPS) is a family of non-negative functions* $\mathbf{m} \rightarrow \Pr(j|\mathbf{m})$ *for* $j = 1, \cdots, J$ *that sum to one.*

Definition 2 *An additive random utility model (ARUM) is a utility field defined by* $\mathbf{U} + \mathbf{m}$*, where* $\mathbf{U} = (U_1, \cdots, U_J)$ *is an absolutely continuously distributed random vector with finite mean, and* $\mathbf{m}$ *is an additive shift vector. The choice probability system associated with this ARUM takes* $\mathbf{m}$ *into probabilities*

$$\Pr(j|\mathbf{m}) = \Pr(U_j + m_j > U_k + m_k \textit{ for all } k \in C \setminus \{j\}) \textit{ for every } j \in C.$$

Let $F(\mathbf{u})$ denote the CDF of $\mathbf{U}$. Then the CDF of the ARUM $\mathbf{U} + \mathbf{m}$ is $F(\mathbf{u} - \mathbf{m})$. The Radon-Nikodym theorem guarantees the existence of a non-negative density $f(\mathbf{u})$ such that

$$F(\mathbf{u}) = \int_{v_J=-\infty}^{u_J} \cdots \int_{v_1=-\infty}^{u_1} f(v_1, \cdots, v_J) dv_1 \cdots dv_J.$$

It also guarantees the existence of the non-negative mixed partial derivatives $\nabla_{\sigma:k} F(\mathbf{u})$ of F. Let F_j denote the derivative of F with respect to u_j. Then the finite mean condition of $\mathbf{U}$ implies

$$\int_{v_j=-\infty}^{\infty} |v_j| F_j(\infty, \cdots, \infty, v_j, \infty, \cdots, \infty) dv_j < \infty. \tag{2.1}$$

For the set S, we define $U_S = \max_{j \in S}(U_j + m_j) - \bar{m}_S$, where $\bar{m}_S = \sum_{j \in S} m_j / s$. The finite mean condition for $\mathbf{U}$ implies the finite mean condition for U_S. The marginal distribution of U_S given $\mathbf{m}_S$ is

$$\begin{aligned} F_{(S)}(u_S|\mathbf{m}_S) &= P\left(U_j + m_j - \bar{m}_S \leq u_S, j \in S\right) \\ &= F(u_S + \bar{m}_S - \mathbf{m}_S, \infty_{-S}). \end{aligned}$$

For a fixed $\mathbf{m}_S$, the distribution of $(U_S, \mathbf{U}_{-S}|\mathbf{m}_S)$ is given by

$$R(u_S, \mathbf{u}_{-S}|\mathbf{m}_S) \equiv F(u_S + \bar{m}_S - \mathbf{m}_S, \mathbf{u}_{-S})$$

and $R(.|\mathbf{m}_S)$ is absolutely continuous.

3 The probability of a single alternative

We begin the analysis by deriving the probability of a single alternative expressed in terms of the probability of another alternative.

The ARUM $\mathbf{U} + \mathbf{m}$ has a choice probability for alternative $j \in C$ that is

$$\begin{aligned} \Pr(j|\mathbf{m}) &= \Pr(U_j + m_j \geq U_k + m_k;\ \ \forall k \in C) \\ &= \int_{-\infty}^{\infty} F_j(u_j + m_j - m_1, \cdots, u_j + m_j - m_J) du_j. \end{aligned} \tag{3.1}$$

Consider the linear transformation $T_j : \mathbf{U} \rightarrow \mathbf{W}$ defined by $W_j = U_j$ and $W_k = U_k - U_j$ for all $k \in C \setminus \{j\}$ with inverse $U_j = W_j$ and $U_k = W_k + W_j$ for all $k \in C \setminus \{j\}$. The determinant of the Jacobian of this transformation is 1. Therefore, the marginal CDF of $\mathbf{W}_{-j}$ under the transformation T_j is given by

$$H_{-j}(\mathbf{w}_{-j}) = \int_{w_j=-\infty}^{\infty} F_j(w_1 + w_j, \cdots, w_{j-1} + w_j, w_j, w_{j+1} + w_j, \cdots, w_J + w_j) dw_j.$$

The choice probability for alternative j is given in terms of H_{-j} by

$$\begin{aligned}\Pr(j|\mathbf{m}) &= \Pr(W_k < m_j - m_k;\ \ \forall k \in C \setminus \{j\}) \\ &= H_{-j}(m_j - \mathbf{m}_{-j}).\end{aligned} \tag{3.2}$$

The above characterization (3.2) of the choice probability in terms of a CDF of utility differences appeared in the early development of ARUM models (McFadden, 1981, 1989), and was used to construct and compute choice probabilities in applications.

For any $i \in C \setminus \{j\}$, we get

$$\begin{aligned}\Pr(i|\mathbf{m}) &= \Pr(W_k < W_i + m_i - m_k;\ \ \forall k \in C \setminus \{i,j\}, \text{and } W_i > m_j - m_i) \\ &= \int\limits_{w_i = m_j - m_i}^{\infty} H_{-j,i}(w_i + m_i - \mathbf{m}_{-j}) dw_i,\end{aligned} \tag{3.3}$$

where $H_{-j,i}(\mathbf{w}_{-j}) = \frac{\partial H_{-j}(\mathbf{w}_{-j})}{\partial w_i}$.

Thus observation of the function $\mathbf{m} \to \Pr(j|\mathbf{m})$ identifies the function H_{-j} through (3.2). Then by (3.3), the probability $\Pr(i|\mathbf{m})$ is identified. We have thus proved

Proposition 1 *In an ARUM, for any $i \neq j \in C$, the function $\mathbf{m} \to \Pr(i|\mathbf{m})$ is identified from observation of the function $\mathbf{m} \to \Pr(j|\mathbf{m})$.*

In the next section, we extend this result up to the knowledge of the probability of any subset of the choice set.

The choice probabilities in (3.2) and (3.3) are entirely and uniquely determined by $H_{-j}(\mathbf{w}_{-j})$ and its derivatives, independent of the density of $W_j = U_j$. Therefore an ARUM $\mathbf{U} + \mathbf{m}$ with $\mathbf{U} \sim F(\mathbf{u})$ has an observationally equivalent ARUM defined for transformation T_j by $\mathbf{W} \sim H_{-j}(\mathbf{w}_{-j})\psi(w_j)$ with $U_j = W_j$ and $U_{-j} = \mathbf{W}_{-j} + W_j$, where ψ_j is any finite mean univariate CDF. This setup, with ψ_j having unit mass at zero and the mapping (3.3), was used by Matzkin (1993) to show that knowledge of a single probability allows one to recover the full CPS.

4 The probability of a set of alternatives

The previous section derived the choice probability for alternative i in terms of the choice probability for a different alternative j. In this section, we repeat this exercise, but with the choice probability of a proper subset S of the choice set instead of alternative j.

The choice probability of the set S is

$$\begin{aligned}\Pr(S|\mathbf{m}) &= \Pr(\max_{j\in S}(U_j + m_j) \geq U_k + m_k;\ \ \forall k \in C) \\ &= \int_{-\infty}^{\infty} R_S(u_S, u_S + \bar{m}_S - \mathbf{m}_{-S}|\mathbf{m}_S) du_S,\end{aligned} \tag{4.1}$$

where $R_S(u_S, \mathbf{u}_{-S}|\mathbf{m}_S) = \partial R(u_S, \mathbf{u}_{-S}|\mathbf{m}_S)/\partial u_S$.

Consider again a linear transformation $T_S : \{U_S, \mathbf{U}_{-S}\} \to W$ defined by $W_S = U_S$ and $W_k = U_k - U_S$ for all $k > s$ with inverse $U_S = W_S$ and $U_k = W_k + W_S$ for all $k > s$. The determinant of the Jacobian of this transformation is 1. Therefore the marginal CDF of $\mathbf{w}_{-S}$ under the transformation T_S is given by

$$H_{-S}(\mathbf{w}_{-S}|\mathbf{m}_S) \equiv \int_{w_S=-\infty}^{\infty} R_S(w_S, \mathbf{w}_{-S} + w_S|\mathbf{m}_S) dw_S.$$

The choice probability for the set S is given under the transformation T_S by

$$\begin{aligned}\Pr(S|\mathbf{m}) &= \Pr(W_k < \bar{m}_S - m_k;\ \ \forall k > s) \\ &= H_{-S}(\bar{m}_S - \mathbf{m}_{-S}|\mathbf{m}_S).\end{aligned} \tag{4.2}$$

For any $i > s$, we get

$$\begin{aligned}\Pr(i|\mathbf{m}) &= \Pr(W_k < W_i + m_i - m_k;\ \ \forall k \in C \setminus \{S \cup i\}, \text{and } W_i > \bar{m}_S - m_i) \\ &= \int_{w_i=\bar{m}_S - m_i}^{\infty} H_{-S,i}(w_i + m_i - \mathbf{m}_{-S}|\mathbf{m}_S) dw_i,\end{aligned} \tag{4.3}$$

where $H_{-S,i}(\mathbf{w}_{-S}|\mathbf{m}_S) = \frac{\partial H_{-S}(\mathbf{w}_{-S}|\mathbf{m}_S)}{\partial w_i}$.

The probability of choice in S, $\Pr(S|\mathbf{m})$, and of the alternatives in the complement of S, $\Pr(i|\mathbf{m})$, have then been expressed in terms of the probability of choice in S. Applying Proposition 1 to a probability in the complement of S shows that the choice probabilities for choice alternatives in S may also be expressed in terms of the probability of choice in S, $\Pr(S|\mathbf{m})$. This establishes the following

Proposition 2 *In an ARUM, for any $i \in C$ and $\emptyset \neq S \subsetneq C$, the function $\mathbf{m} \to \Pr(i|\mathbf{m})$ is identified from observation of the function $\mathbf{m} \to \Pr(S|\mathbf{m})$.*

5 Limits to identification

It is clear that adding a univariate random variable to all random utilities in an ARUM does not affect choice probabilities. This means that the random utility vector $\mathbf{U}$ is not identified from observation of discrete choices. The following proposition makes this insight a little more precise by establishing a converse, namely that if two ARUM yield identical choice probabilities then their random utility vectors must be identical up to an additive random variable. This is then the limit for identification in an ARUM: it is possible to identify the distribution of $\mathbf{U}$ up to an additive univariate random variable and not more.

Proposition 3 *Two ARUM $\mathbf{m}+\mathbf{U}$ and $\mathbf{m}+\mathbf{U}'$ yield the same choice probabilities for all $\mathbf{m}$ if and only if there exists independent univariate random variables $\delta, \delta' \in \mathbb{R}$, such that $\mathbf{U}+\delta$ has the same distribution as $\mathbf{U}'+\delta'$.*

Proof of Proposition 3. An ARUM $\mathbf{U}+\mathbf{m}$ yields the same choice probabilities as an ARUM $\mathbf{U}+\mathbf{m}+\delta$, where $\delta \in \mathbb{R}$ is a univariate random variable that is added to the indirect utility of all alternatives.

Conversely, if two ARUM $\mathbf{U}+\mathbf{m}$ and $\mathbf{U}'+\mathbf{m}$ yield the same choice probabilities, then the same is true of $(0, U_2-U_1, ..., U_J-U_1)+\mathbf{m}$ and $(0, U_2'-U_1', ..., U_J'-U_1')+\mathbf{m}$. The probability that alternative 1 is chosen in the first model is

$$\Pr(1|\mathbf{m}) = \Pr(U_2-U_1 \leq m_1-m_2, ..., U_J-U_1 \leq m_1-m_J),$$

a similar expression applies in the second model, and this shows that the CDF of $(U_2-U_1, \cdots, U_J-U_1)$ and $(U_2'-U_1', ..., U_J'-U_1')$ are the same. Defining $\delta=-U_1$ and $\delta'=-U_1'$ completes the proof. ■

Let $\psi:\mathbb{R}\to\mathbb{R}$ be an increasing function. Then a RUM with utilities $(\psi(U_1+m_1), ...\psi(U_J+m_J))$ yields the same choice probabilities as the ARUM $\mathbf{U}+\mathbf{m}$ and is hence observationally equivalent to the ARUM. This topic is explored in Mattsson et al. (2014).

6 The expected maximum utility

In this section, we establish a proposition that indicates a way in which an ARUM can be constructed with a nested structure that corresponds to a partitioning of choice alternatives into groups. To do this, we rely on the properties of the expected maximum utility of an ARUM. The random maximum utility is denoted $Y \equiv \max_{j\in C}(U_j+m_j)$ and it has a CDF that is $F(y-$

$m_1, \cdots, y - m_J)$. The expected maximum utility is denoted

$$h(\mathbf{m}) \equiv \mathbb{E}(Y) = \mathbb{E}(\max_{j \in C}(U_j + m_j)) = \int_{-\infty}^{\infty} y[\frac{d}{dy} F(y - m_1, \cdots, y - m_J)]dy. \tag{6.1}$$

The expected maximum utility is of interest in its own right as a measure of the welfare of the decision-maker. It is also useful as it encodes the choice probabilities of the ARUM in a single function. This is stated in the following well-known result (see e.g. Fosgerau et al., 2013).

Proposition 4 *Let $h(\mathbf{m})$ be the expected maximum utility of an ARUM $\mathbf{U} + \mathbf{m}$ as defined in (6.1). Then*

$$h(\mathbf{m}) = \int_{0}^{\infty} [1 - F(y - \mathbf{m})]dy - \int_{-\infty}^{0} F(y - \mathbf{m})dy. \tag{6.2}$$

and differentiating with respect to m_j, we get

$$h_j(\mathbf{m}) \equiv \partial h(\mathbf{m})/\partial m_j = \Pr(j|\mathbf{m}). \tag{6.3}$$

We are then able to state the result of this section.

Proposition 5 *Let $h : \mathbb{R}^J \to \mathbb{R}$ and $h^k : \mathbb{R}^{J_k} \to \mathbb{R}, k = 1, ..., J$ be expected maximum utilities of ARUM. Then $h\left(h^1, ..., h^J\right)$ is also the expected maximum utility of an ARUM.*

Proof of Proposition 5. Let $h : \mathbb{R}^J \to \mathbb{R}$ and $h' : \mathbb{R}^{J'} \to \mathbb{R}$ be expected maximum utilities of ARUM. We will first show that

$$h\left(h'\left(\mathbf{m}'\right), \mathbf{m}_{-1}\right) : \mathbb{R}^{J'+J-1} \to \mathbb{R}$$

is also a CPGF. Let h correspond to the ARUM $\mathbf{m} + \mathbf{U}$ and h' to the ARUM $\mathbf{m}' + \mathbf{U}'$, where $\mathbf{U}$ and $\mathbf{U}'$ are independent. Consider an ARUM with $J + J' - 1$ alternatives indexed by jj' where $j \in \{1, ..., J\}$ and $j' \in \{1, .., J'\}$ when $j = 1$ and $j' = 1$ otherwise. Define $U''_{jj'} = U_j + U'_{j'} 1_{\{j=1\}}$ and consider the ARUM $m''_{jj'} + U''_{jj'}$ with associated CPGF denoted $h''\left(\mathbf{m}''\right)$.

Then

$$\begin{aligned}
h''(\mathbf{m}'') &= \mathbb{E}\max_{jj'}\left\{m''_{jj'}+U''_{jj'}\right\} \\
&= \mathbb{E}\max_{jj'}\left\{m''_{jj'}+U_j+U'_{j'}1_{\{j=1\}}\right\} \\
&= \mathbb{E}\left(\mathbb{E}\left(\max_{jj'}\left\{m''_{jj'}+U_j+U'_{j'}1_{\{j=1\}}\right\}|U\right)\right) \\
&= \mathbb{E}\left(\max\left\{U_1+\mathbb{E}\max_{j'}\left\{m''_{1j'}+U'_{j'}\right\},\max_{j>1}\left\{m''_{j1}+U_j\right\}\right\}\right) \\
&= \mathbb{E}\left(\max\left\{U_1+h'\left(m''_{11},...,m''_{1J'}\right),\max_{j>1}\left\{m''_{j1}+U_j\right\}\right\}\right) \\
&= h\left(h'\left(m''_{11},...,m''_{1J'}\right),m''_{21},...,m''_{J1}\right).
\end{aligned}$$

Proposition 5 follows from repeated application of this result. ■

From the proof of Proposition 5 we observe that if an ARUM has an expected maximum utility of the form $h\left(h^1,...,h^J\right)$, then it has a random utility representative of the form $\mathbf{m}+\mathbf{U}$ where $U_{jj'}=\delta_j+\eta_{j'}$, where the δ_j are random utilities corresponding to h, and $\eta_{j'}$ are random utilities corresponding to each nest defined by h^i. The random utility components $\eta_{j'}$ are independent between nests and independent of δ.

Example 1 *We will use Proposition 5 to construct a two-level nested logit model. Let $\mu>0$ and define*

$$h(\mathbf{m})=\mu\ln\left(\sum_{k=1}^{J}e^{m_k/\mu}\right),\mathbf{m}\in\mathbb{R}^J$$

as the expected maximum utility of a multinomial logit model. Define similarly

$$h^k\left(\mathbf{m}^k\right)=\mu_k\ln\left(\sum_{j=1}^{J_k}e^{m_j^k/\mu_k}\right),\mathbf{m}^k\in\mathbb{R}^{J_k}$$

for $\mu_k>0$. Then by Proposition 5, the composition

$$h\left(h^1\left(\mathbf{m}^1\right),...,h^J\left(\mathbf{m}^J\right)\right)=\mu\ln\left(\sum_{k=1}^{J}\left(\sum_{j=1}^{J_k}e^{m_j^k/\mu_k}\right)^{\frac{\mu_k}{\mu}}\right)$$

is the maximum expected utility of an ARUM. According to Proposition 4, choice probabilities can be found by differentiation,

$$\frac{\partial\left(h\left(h^1\left(\mathbf{m}^1\right),...,h^J\left(\mathbf{m}^J\right)\right)\right)}{\partial m_i^k}=\frac{e^{m_i^k/\mu_k}}{\sum_{j=1}^{J}e^{m_j^k/\mu_k}}\frac{\exp\left(\frac{\mu_k}{\mu}\ln\left(\sum_{j=1}^{J_k}e^{m_j^k/\mu_k}\right)\right)}{\sum_{k'=1}^{J}\exp\left(\frac{\mu_{k'}}{\mu}\ln\left(\sum_{j=1}^{J_{k'}}e^{m_j^{k'}/\mu_{k'}}\right)\right)},$$

and these may be recognized as belonging to a nested logit model (McFadden, 1978).

7 Conclusion

This note has shown that all the observational content of an ARUM is encoded in the probability of a group of alternatives. As observed by McFadden (2014) (for the case when the probability of a single alternative is observed), the good news is that this reduces the requirements on data in empirical applications of ARUM. The bad news is that it is not possible to specify an ARUM in a partial way, where only the probability of a single or a group of alternatives is specified, without implicitly specifying the whole model. Furthermore, we have shown that further identification cannot be obtained and we have presented a general way to construct ARUM with a built-in partitioning of alternatives into groups.

The present results rely crucially on the defining property of ARUM that a vector of random utilities is perturbed by additive, non-random location shifts. There is a small but important literature on identification in more general discrete choice models. A recent point of entry to that literature is provided by Berry and Haile (2016).

References

Amemiya, T. (1981) Qualitative Response Models: A Survey *Journal of Economic Literature* **19**(4), 1483–1536.

Berry, S. and Haile, P. (2016) Identification in Differentiated Products Markets *Annual Review of Economics* **8**(1), 27–52.

Fosgerau, M., McFadden, D. and Bierlaire, M. (2013) Choice probability generating functions *Journal of Choice Modelling* **8**, 1–18.

Maddala, G. (1983) *Limited-Dependent and Qualitative Variables in Econometrics* Cambridge University Press Cambridge.

Mattsson, L.-G., Weibull, J. W. and Lindberg, P. O. (2014) Extreme values, invariance and choice probabilities *Transportation Research Part B: Methodological* **59**, 81–95.

Matzkin, R. L. (1993) Nonparametric identification and estimation of polychotomous choice models *Journal of Econometrics* **58**(1-2), 137–168.

Matzkin, R. L. (2007) Chapter 73 Nonparametric identification *in* J. J. H. a. E. E. Leamer (ed.), *Handbook of Econometrics* Vol. 6, Part B Elsevier pp. 5307–5368.

McFadden, D. (1974) Conditional Logit Analysis of Qualitative Choice Behaviour *Frontiers in Econometrics* Academic Press New York pp. 105–142.

McFadden, D. (1978) Modelling the choice of residential location *in* A. Karlquist, F. Snickars and J. W. Weibull (eds), *Spatial Interaction Theory and Planning Models* North Holland Amsterdam pp. 75 –96.

McFadden, D. (1981) Econometric Models of Probabilistic Choice *in* C. Manski and D. McFadden (eds), *Structural Analysis of Discrete Data with Econometric Applications* MIT Press Cambridge, MA, USA pp. 198–272.

McFadden, D. (1989) A method of simulated moments for estimation of the multinomial Probit without numerical integration *Econometrica* **57**(5), 995–1026.

McFadden, D. L. (2014) CPGF construction: A refinement of Fosgerau, McFadden, and Bierlaire (2013) "Choice Probability Generating Functions", Journal of Choice Modeling, 8, 1-18.