Choice probability generating functions

Fosgerau, Mogens and McFadden, Daniel and Bierlaire, Michel

Technical University of Denmark, University of California at Berkeley, EPFL

2010

Choice probability generating functions

Mogens Fosgerau* Daniel McFadden† Michel Bierlaire‡

June 4, 2010

Abstract

This paper establishes that every random utility discrete choice model (RUM) has a representation that can be characterized by a choice-probability generating function (CPGF) with specific properties, and that every function with these specific properties is consistent with a RUM. The choice probabilities from the RUM are obtained from the gradient of the CPGF. Mixtures of RUM are characterized by logarithmic mixtures of their associated CPGF. The paper relates CPGF to multivariate extreme value distributions, and reviews and extends methods for constructing generating functions for applications. The choice probabilities of any ARUM may be approximated by a cross-nested logit model. The results for ARUM are extended to competing risk survival models.

*Technical University of Denmark, Centre for Transport Studies, Sweden and Ecole Nationale Supèrieure. mf@transport.dtu.dk

†Econometrics Laboratory, University of California, Berkeley

‡Transport and Mobility Laboratory, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland, michel.bierlaire@epfl.ch

1 Introduction

This paper considers decision-makers who make discrete choices that are draws from a multinomial choice probability vector indexed by the alternatives in a finite choice set. A random utility model (RUM) associates a vector of random (from the analyst's viewpoint) utilities with the choice set alternatives, and postulates that the choices of decision-makers maximize utility. If the distribution of the utility vectors is absolutely continuous, then it induces a unique multinomial choice probability vector. This paper is concerned with determining necessary and sufficient conditions for a vector of choice probabilities to be consistent with a RUM. This question is answered in conventional revealed preference theory (Samuelson, 1947; Houthakker, 1950; Richter, 1966) for choice sets that are economic budget sets, and in the theory of stochastic revealed preference for finite choice sets (McFadden and Richter, 1990; McFadden, 2005). This paper adds to this literature by showing that a multinomial choice probability vector is consistent with a RUM if and only if it is the gradient of a choice probability generating function (CPGF) with specified properties that can be checked in applications. This is the analogue for discrete choice and the random utility model of the Antonelli (1886) conditions in conventional economic demand analysis; see, for example, Mas-Colell et al. (1995, 3H).

This paper embeds RUM in additive random utility model (ARUM) families, obtained by adding location shift vectors to the random utility vector of the original RUM. The CPGF is given from the expected maximum utility of ARUM family members, considered as a function of the location shifts. Then the CPGF is a transform of a multivariate probability distribution that is related to a multivariate Laplace transform, and retains exactly those aspects of the distribution of utilities that are relevant for choice probabilities.

This paper goes on to define the exponent (EXP) of a multivariate CDF, and shows that the EXP has the properties of a CPGF if and only if the distribution of utilities is multivariate extreme value (MEV). Rules are given that allow exponents to be combined into new exponents. This allows the combination of MEV distributions into new MEV distributions. The associated copulas may be viewed as generalizations of the Archimedian copula.

The mapping from MEV distributions to EXP is many-to-one, and MEV distributions that generate observationally distinct ARUM, and hence distinct CPGF, can have the same EXP. Therefore, even though the EXP of a specific MEV distribution has the properties of a CPGF, it may not be the CPGF associated with this particular MEV distribution. However, if a MEV that generates a particular EXP satisfies an additional condition termed cross-alternative max-stability (CAMS), then this MEV has marginal distributions that are extreme value type 1 distributed, and a CPGF that coincides with the EXP. Put another way, if EXP is the exponent of a MEV distribution, then there is another MEV distribution, not necessarily observationally equivalent, that has extreme value type 1 marginals and the same exponent EXP, and EXP is its CPGF. The family of CAMS MEV induce ARUM that include the multinomial logit model, as well as generalizations such as the nested logit model and the cross-nested logit model. This paper shows that the choice probabilities of every ARUM may be approximated by a cross-nested logit model. This shows that even though the set of ARUM from CAMS MEV is a strict subset of all ARUM, it is in a sense dense in the set of all ARUM.

Multiple risk survival models are similar to RUM, with duration and cause of exit playing roles analogous to utility and choice. A difference is that durations are positive and the smallest duration is observed, while utility has no sign restriction and is not observed. The paper extends to the case of multiple risk survival models the main results relating CPGF and MEV to RUM.

Section 2 discusses general RUM. Section 3 considers the special case of RUM based on MEV distributions. Section 4 extends some results to multiple risk survival models. Section 5 concludes. Table 1 summarizes the acronyms used in this paper.

Acronym	Definition
RUM	Random Utility Model
ARUM	Additive Random Utility Model (additive location shifts)
CPGF	Choice Probability Generating Function
EXP	Exponent of a multivariate distribution
MEV	Multivariate Extreme Value distribution
CAMS	Cross-Alternative Max-Stability
TARUM	Transformed Additive Random Utility Model
ASP	Alternating signs property

Table 1: Acronyms

2 Random utility models

Suppose decision-makers face a finite set of alternatives indexed by a choice set $\mathcal{C} = \{1, \dots, J\}$. Decision-makers and alternatives may be described by explanatory variables, and the dependence of utilities on these variables may have structure derived from assumptions of economic rationality. We suppress these for the current discussion, and reintroduce them later. We also initially suppress notation for mixing of RUM.

Notation Let $\sigma = (\sigma(1), \dots, \sigma(J))$ denote a permutation of $(1, ..., J)$, and let $\sigma : k = (\sigma(1), \dots, \sigma(k))$ denote the first k elements of σ. Let

$$\nabla_{\sigma:k} F(u) \equiv \frac{\partial^k F(u_1, \dots, u_J)}{\partial u_{\sigma(1)} \cdots \partial u_{\sigma(k)}}$$

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

$$\nabla_{1\ldots J} F(u) \equiv \frac{\partial^J F(u_1, \ldots, u_J)}{\partial u_1 \cdots \partial u_J}.$$

Definition 1 *A* random utility model *(RUM) is a vector $U = (U_1, \ldots, U_J)$ of latent random variables in $\mathbb{R}^J$, with a Cumulative Distribution Function (CDF) $R(u_1, \ldots, u_J)$ that is absolutely continuous with respect to Lebesgue measure, and has a density given by $\nabla_{1\ldots J} R(u_1, \ldots, u_J)$. A RUM induces an observable choice probability*

$$(1) \qquad P_{\mathcal{C}}(j) = \Pr(U_j > U_k \textit{ for } j \neq k \in \mathcal{C}) = \int_{-\infty}^{+\infty} \nabla_j R(u, \ldots, u) \, \mathrm{d}u,$$

where

$$\nabla_j R(u, \ldots, u) \equiv \frac{\partial R}{\partial u_j}(u, \ldots, u).$$

Remark The definition of a RUM ensures that for any σ and $k = 1, ..., J$, the derivative $\nabla_{\sigma:k} R(u_1, \ldots, u_J)$ exists almost everywhere and is non-negative. It ensures for each j that the marginal CDF $R_{(j)}(u_j) = R(+\infty, \ldots, +\infty, u_j, +\infty, \ldots, +\infty)$ is absolutely continuous on $\mathbb{R}$. It also ensures that the probability of ties is zero; i.e. $\sum_{j=1}^J P_{\mathcal{C}}(j) = 1$. The RUM definition does not exclude zero probability alternatives.

2.1 Embedding and observational equivalence

For any continuously differentiable increasing transformation $r : \mathbb{R} \to \mathbb{R}$, which may be stochastic, the image $r(U) \equiv (r(U_1), \ldots, r(U_J))$ of a RUM $U = (U_1, \ldots, U_J)$ is again a RUM with the same associated choice probability. Then the family of such increasing transformations of a RUM defines an observationally equivalent class. A representative can always be chosen from this equivalence class that has $E|U_j|$ finite for all j. For example,

$$U_j^* = 1/(1 + exp(-U_j)) \tag{2}$$

has this property. Then, we will consider without loss of generality only RUM representations that have finite means.

Definition 2 *If $U^0 = (U_1^0, \ldots, U_J^0)$ is a finite-mean RUM with CDF $R(u_1, \ldots, u_J)$, and choice probability $P_{\mathcal{C}}^0(j)$, and $V = (V_1, \ldots, V_J) \in \mathbb{R}^J$ is a location vector, then*

$$U = (U_1, \ldots, U_J) = V + U^0 \tag{3}$$

is again a finite-mean RUM with CDF $R(u_1 - V_1, \ldots, u_J - V_J)$, and choice probability

$$P_{\mathcal{C}}(j|V) = \Pr(U_j \geq U_k \mathit{\ for\ } k \in \mathcal{C}) = \Pr(U_j^0 + V_j \geq U_k^0 + V_k \mathit{\ for\ } k \in \mathcal{C}). \tag{4}$$

A family of RUM with finite means indexed by location vectors V is termed an additive random utility model *(ARUM).*

Remark Obviously every RUM can be embedded in a ARUM, and recovered by setting $V = 0$. Then, without loss of generality, we can represent a RUM as an ARUM. Note that ARUM have been defined in previous literature as forms $U_j = V_j + U_j^0$ that are the sum of a systematic component of utility V_j and a disturbance U_j^0. Our definition has the same mathematical structure, so that we retain the name for this family, but we interpret the V_j as location shifts from a base RUM U^0, dissociated from the utility.

Remark If a random vector $W = (W_1, \ldots, W_J)$ has an absolutely continuous CDF $R(w_1, \ldots, w_J)$ with marginal CDF's $R_{(j)}(w_j)$, and $z_j = s_j(w_j)$ is a continuously differentiable increasing transformation with a (also continuously differentiable) inverse $w_j = S_j(z_j)$, then $Z = (Z_1, \ldots, Z_J) = (s_1(W_1), \ldots, s_J(W_J))$ has a CDF $F(z_1,\dots,z_J) = R(S_1(z_1),\dots,S_J(z_J))$, and conversely $R(w_1,\dots,w_J) = F(s_1(w_1),\dots,s_J(w_J))$. The associated densities are related by

$$(5) \qquad \nabla_{1\dots J}R(w_1,\dots,w_J) = \nabla_{1\dots J}F(s_1(w_1),\dots,s_J(w_J))\prod_{j=1}^{J} s_j'(w_j),$$

where the product in this expression is the Jacobian of the nonlinear transformation. In particular, the transformation $Z_j = r_j(W_j) \equiv \psi^{-1}(R_{(j)}(W_j))$, where ψ is a continuously differentiable increasing univariate CDF whose support is a bounded or unbounded interval $\Psi \in \mathbb{R}$, defines a random vector $Z \in \Psi^J$ with a CDF $F(z_1,\dots,z_J)$ whose univariate marginals are $\psi(z_j)$ for each j. The transformation $z_j = s_j(w_j)$ can in general depend on observed and stochastic variables, and unknown parameters, and when this is the case, the probability statements above are interpreted as conditioned on these factors.

Definition 3 *A finite-mean family of RUM defined implicitly by $\varepsilon_j = r_j(U_j - V_j)$, where r_j is a continuously differentiable increasing transformation with a continuously differentiable inverse, $\varepsilon = (\varepsilon_1,\dots,\varepsilon_J)$ has an absolutely continuous CDF $F(\varepsilon)$, and $V = (V_1,\dots,V_J)$ is a location vector, is termed a finite-mean* transformed additive random utility model *(TARUM).*

Define $\zeta_j \equiv r_j^{-1}(\varepsilon_j)$ and $\zeta = (\zeta_1,\dots,\zeta_J)$. Then $U_j = V_j + \zeta_j$ is a finite-mean ARUM with a CDF $R(U-V) \equiv F(r(U-V))$ that is absolutely continuous, with $r(U-V) = (r_1(U_1-V_1),\dots,r_J(U_J-V_J))$, and this ARUM is observationally equivalent to the TARUM.

Lemma 4 *The choice probability associated with a TARUM satisfies*

(6)
$$\begin{aligned} P_{\mathcal{C}}(j) &= \Pr(U_j > U_k \textit{ for } j \neq k \in \mathcal{C}) \\ &= \Pr(r_k(U_j - V_k) > r_k(U_k - V_k) \textit{ for } j \neq k \in \mathcal{C}) \\ &= \int_{-\infty}^{+\infty} \nabla_j F(r_1(u - V_1), \ldots, r_J(u - V_j)) r_j'(u - V_j)\, \mathrm{d}u \\ &\equiv \int_{-\infty}^{+\infty} \nabla_j R(u - V_1, \ldots, u - V_J)\, \mathrm{d}u. \end{aligned}$$

Proof. Given the absolutely continuous CDF $F(\varepsilon)$ of $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_J)$ with density $\nabla_{1\ldots J} F(\varepsilon)$, the density of U is $\nabla_{1\ldots J} F(r(u - V)) \prod_{j=1}^{J} r_j'(u_j - V_j)$. Then

$$\begin{aligned} P_{\mathcal{C}}(j) &= \Pr(U_j > U_k \text{ for } j \neq k \in \mathcal{C}) \\ &= \int_{u_j=-\infty}^{+\infty} \int_{u_1=-\infty}^{u_j} \cdots \int_{u_{j-1}=-\infty}^{u_j} \int_{u_{j+1}=-\infty}^{u_j} \cdots \int_{u_J=-\infty}^{u_j} \\ &\qquad \nabla_{1\ldots J} F(r(u - V)) \prod_{j=1}^{J} r_j'(u_j - V_j)\, \mathrm{d}u_1 \ldots \mathrm{d}u_J \\ &= \int_{u_j=-\infty}^{+\infty} \nabla_j F(r_1(u_j - V_1), \ldots, r_J(u_j - V_J)) r_j'(u_j - V_j)\, \mathrm{d}u_j. \end{aligned}$$

By definition, $R(\zeta) \equiv F(r(\zeta))$, so that $\nabla_j R(\zeta) \equiv \nabla_j F(r(\zeta)) r_j'(\zeta_j)$. Then,

$$\int_{u_j=-\infty}^{+\infty} \nabla_j F(r_1(u_j - V_1), \ldots, r_J(u_j - V_J)) r_j'(u_j - V_j)\, \mathrm{d}u_j = \int_{-\infty}^{+\infty} \nabla_j R(u - V_1, \ldots, u - V_J)\, \mathrm{d}u.$$

■

Remark Requiring that a representation of a RUM have finite mean implies that the random vector ζ has a finite mean; this does not require that ε has a finite mean. The primary interest in TARUM is that the nonlinear transformations r_j can absorb dependence of stochastic terms on explanatory variables and the effects of heteroskedasticity, so that ε is clear of these application-specific effects. Suppose an ARUM with an absolutely continuous CDF $R(u_1 - V_1, \ldots, u_J - V_J)$ and marginal CDF's $R_{(j)}$, and define $\varepsilon_j = r_j(U_j - V_j) \equiv \psi^{-1}(R_{(j)}(U_j - V_j))$, where ψ is a continuous increasing univariate CDF. Then $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_J)$ has a CDF $F(\varepsilon)$ with specified univariate marginals ψ. TARUM include Box-Cox transformations such as

$$r(U_i - V_i) = \left(1 - e^{-\lambda(U_i - V_i)}\right)/\lambda,$$

polynomials such as

$$r(U_i - V_i) = \sum_{k=1}^{K} a_k (U_i - V_i)^{2k-1}$$

with $a_k \geq 0$, and other transformations of location and scale, and r can also be a function of observed explanatory variables and latent stochastic vectors that capture taste heterogeneities.

Remark In a TARUM $\varepsilon_j = r_j(U_j - V_j)$ where $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_J)$ has a CDF $F(\varepsilon)$, choice probabilities $P_{\mathcal{D}}(j|V)$ from a non-empty subset $\mathcal{D}$ of $\mathcal{C}$ are obtained by first forming the marginal distribution of the components of ε associated with alternatives in $\mathcal{D}$, and then applying (6) using this distribution. Equivalently, $P_{\mathcal{D}}(j|V)$ is obtained from $P_{\mathcal{C}}(j|V)$ by setting $V_k = -\infty$ for $k \notin \mathcal{D}$.

Together, the definitions and remarks above establish that every RUM has a representation that can be embedded in an ARUM with finite mean, or an observationally equivalent TARUM whose univariate marginals are a specified continuous increasing univariate CDF ψ.

2.2 Choice Probability Generating Functions

This section defines a class of choice-probability generating functions (CPGF) with specific properties. The main result shows that every TARUM has an associated CPGF whose first derivatives are the choice probabilities of the TARUM, and conversely that every function with the specific properties is the CPGF associated with an ARUM. Thus, an observationally equivalent family of RUM are fully characterized by an associated CPGF.

Definition 5 *A function g on $[0, +\infty)^J$ has the* Alternating Signs Property *(ASP) if for any permutation σ of $(1, \ldots, J)$ and $k = 1, \ldots, J$, the signed mixed derivatives*

$$(-1)^{k-1} \nabla_{\sigma:k} g(y_1, \ldots, y_J)$$

exist, are independent of the order of differentiation, and are non-negative.

Definition 6 *A* choice-probability generating function *(CPGF) is an extended non-negative function $G : [0, +\infty]^J \to [0, +\infty]$ with the properties:*

[G1] *(Weak ASP) $\ln G(y)$ satisfies the ASP, so that for any permutation σ of $(1, \ldots, J)$ and $k = 1, \ldots, J$, its mixed partial derivatives are independent of the order of differentiation and*

$$\chi_{\sigma:k}(\ln y) \equiv (-1)^{k-1} \nabla_{\sigma:k} \ln G(y) \geq 0. \tag{7}$$

[G2] *(Homogeneity) For each $\lambda > 0$ and $y \in [0, +\infty]^J$, $G(\lambda y) = \lambda G(y)$.*

[G3] *(Boundary) $G(0) = 0$, and for $j = 1, \ldots, J$, if $1_{(j)}$ denotes a unit vector with the jth component equal to one, then*

$$\lim_{\lambda \to \infty} G(\lambda 1_{(j)}) \to \infty.$$

[G4] *(Integrability) Let* $V = (V_1, \ldots, V_J)$, $e^V = \left(e^{V_1}, \ldots, e^{V_J}\right)$, $L = \{V \in \mathbb{R}^J | \sum_{j=1}^J V_j = 0\}$. *Then the following holds:*

$$\int_L \chi_{1,\ldots,J}(V)dV = J^{-1},$$

$$\int_L |V_j|\chi_{1,\ldots,J}(V)dV < +\infty, \textit{ for } j = 1, \ldots, J,$$

$$\max_{k \neq j} V_k - V_j \to \infty \textit{ implies } \chi_j(V) \to 0.$$

Remark The gradient $P_{\mathcal{C}}(j|V) = \partial \ln G(e^V)/\partial V_j \equiv \chi_j(V)$ is the choice probability generated by the CPGF G, with [G1] and [G2] ensuring that the probability is non-negative and sums to one. An implication of [G2] is that

$$\chi_{\sigma:k}(V - c) \equiv \chi_{\sigma:k}(V)$$

for any scalar c.

Remark If $\mathcal{D}$ is a non-empty subset of $\mathcal{C}$, then its CPGF is obtained from G by setting its arguments to zero for alternatives that are not in $\mathcal{D}$.

Remark If G is a CPGF, then from [G1],

$$\partial^2 \ln G(e^V)/\partial V_j \partial V_k \leq 0 \text{ for } j \neq k. \tag{8}$$

From [G2] and Euler's theorem, one has $\sum_{j=1}^J \partial \ln G(e^V)/\partial V_j \equiv 1$, implying

$$\partial P_{\mathcal{C}}(j|V)/\partial V_j = \partial^2 \ln G(e^V)/\partial V_j^2 = -\sum_{k \neq j} \partial^2 \ln G(e^V)/\partial V_j \partial V_k \geq 0. \tag{9}$$

Then, the hessian of $\ln G(e^V)$ has a *weakly dominant positive diagonal*, implying that $\ln G(e^V)$ is a convex function of V.

Examples of CPGF The linear function

$$G(e^V) = \sum_{j=1}^J e^{V_j}, \tag{10}$$

satisfies [G1]–[G4] and generates the logit model. The function

$$G(e^V) = \sum_{j=1}^{J} \sum_{k>j} \alpha_{jk} \left(e^{\mu_{jk} V_j} + e^{\mu_{jk} V_k}\right)^{1/\mu_{jk}}, \tag{11}$$

with $\alpha_{jk} \geq 0$ and $\mu_{jk} \geq 1$ satisfies [G1]–[G4] and generates cross-nested logit models. However, if $\mu_{jk} < 1$ for some j, k in the last function, then G fails to satisfy [G1].

Theorem 7 *If $\varepsilon_j = r_j(U_j - V_j)$ is a finite-mean TARUM in which $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_J)$ has CDF $F(\varepsilon)$ and $U = (U_1, \ldots, U_J)$ has finite mean, and an observationally equivalent ARUM is $U_j = V_j + \zeta_j$ with $\zeta = (\zeta_1, \ldots, \zeta_J)$ having CDF $R(\zeta) \equiv F(r(U - V))$, where $r(U - V) = (r_1(U_1 - V_1), \ldots, r_J(U_J - V_J))$, then an associated CPGF $G(e^V)$ defined by*

$$\begin{aligned} \ln G(e^V) &\equiv E[\max\nolimits_{j\in\mathcal{C}} U_j] \\ &\equiv \int_0^{+\infty} [1 - R(u - V)]du - \int_{-\infty}^{0} R(u - V)du \end{aligned} \tag{12}$$

exists and satisfies [G1]-[G4]. The choice probability implied by the TARUM satisfies

$$P_{\mathcal{C}}(j|V) = \Pr(U_j > U_k \textit{ for } k \neq j) = \partial \ln G(e^V)/\partial V_j. \tag{13}$$

Conversely, if G is a function satisfying properties [G1]-[G4], then there exists an ARUM such that G is an associated CPGF and (13) holds.

Proof. Consider first sufficiency. Assume that $\varepsilon_j = r_j(U_j - V_j)$ is a TARUM and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_J)$ has CDF $F(\varepsilon)$. Let $U_j = V_j + \zeta_j$ with $\zeta = (\zeta_1, \ldots, \zeta_J)$ having CDF $R(\zeta)$ be an observationally equivalent ARUM.

Consider [G1]: The requirement that the TARUM have finite mean utility implies that $U_{\mathcal{C}} = \max_{j\in\mathcal{C}} U_j$ has $E|U_{\mathcal{C}}| \leq \sum_{j=1}^{J} E|U_j| < \infty$. Then $\ln G(e^V) \equiv$

$E[U_{\mathcal{C}}]$ exists. Then, integrating by parts,

(14)
$$\begin{aligned}\ln G(e^V) \equiv E[U_{\mathcal{C}}|V] &= \int_{-\infty}^{+\infty} u[dR(u-V)/du]du \\ &= \int_{0}^{+\infty} [1-R(u-V)]du - \int_{-\infty}^{0} R(u-V)du,\end{aligned}$$

and (12) holds. Differentiating (14) with respect to V_j yields

$$\partial \ln G(e^V)/\partial V_j = \int_{-\infty}^{+\infty} \nabla_j R(u-V_1,\ldots,u-V_J)du, \tag{15}$$

and from (6), the result (13) holds.

Mixed differentiation of $\ln G$ for any permutation σ of $(1,\ldots,J)$ and $k = 1,\ldots,J$ gives

(16)
$$\begin{aligned}\chi_{\sigma:k}(V) &= (-1)^{k-1}\partial^k \ln G(e^V)/\partial V_{\sigma(1)}\ldots\partial V_{\sigma(k)} \\ &= \int_{-\infty}^{+\infty} \nabla_{\sigma:k} R(u-V_1,\ldots,u-V_J)du \geq 0.\end{aligned}$$

Hence, [G1] holds.

Consider [G2]: If $\lambda > 0$, then

(17)
$$\begin{aligned}\ln G(\lambda e^V) &= \int_{0}^{+\infty} [1-R(u-\ln\lambda-V)]du - \int_{-\infty}^{0} R(u-\ln\lambda-V)du \\ &= \int_{-\ln\lambda}^{+\infty} [1-R(u-V)]du - \int_{-\infty}^{-\ln\lambda} R(u-V)du \\ &= \ln\lambda + \int_{0}^{+\infty} [1-R(u-V)]du - \int_{-\infty}^{0} R(u-V)du \\ &= \ln\lambda + \ln G(e^V),\end{aligned}$$

implying that $G(y)$ is homogeneous of degree one on $y \equiv e^V \geq 0$ and [G2] holds. Letting $\lambda = \exp(-c)$ for any scalar c, (17) implies

$$\chi_j(V) \equiv \chi_j(V - c).$$

Consider [G3]. From (14),

$$\ln G(1_{(j)}) = \int_0^{+\infty} [1 - R_{(j)}(u)]du - \int_{-\infty}^0 R_{(j)}(u)du = E[\zeta_j]$$

is finite, so that $G(1_{(j)}) > 0$, implying with [G2] that

$$\lim_{\lambda \to \infty} G(\lambda 1_{(j)}) \to \infty.$$

[G2] also implies $G(0) = 0$. Hence, [G3] holds.

Consider [G4]. First, suppose

$$w = \max_{k>1} V_k - V_1 \to \infty.$$

Then,

(18)

$$\begin{aligned}
\chi_1(V) &= \int_{u=-\infty}^{\sqrt{w}} \nabla_1 R(u, u - V_2 + V_1, \ldots, u - V_J + V_1)\,\mathrm{d}u \\
&\qquad + \int_{u=\sqrt{w}}^{+\infty} \nabla_1 R(u, u - V_2 + V_1, \ldots, u - V_J + V_1)\,\mathrm{d}u \\
&\leq \int_{u=-\infty}^{\sqrt{w}} \nabla_1 R(u, \sqrt{w} - V_2 + V_1, \ldots, \sqrt{w} - V_J + V_1)\,\mathrm{d}u \\
&\qquad + \int_{u=\sqrt{w}}^{+\infty} \nabla_1 R(u, +\infty, \ldots, \infty)\,\mathrm{d}u \\
&\leq R(\sqrt{w}, \sqrt{w} - V_2 + V_1, \ldots, \sqrt{w} - V_J + V_1) + [1 - R(\sqrt{w}, +\infty, \ldots, +\infty)] \\
&\leq R(+\infty, \ldots, +\infty, \sqrt{w} - w, +\infty, \ldots, +\infty) + [1 - R(\sqrt{w}, +\infty, \ldots, +\infty)] \\
&\to 0 \text{ as } w \to +\infty.
\end{aligned}$$

Since this argument holds for any permutation of $(1, \dots, J)$, the last property in [G4] holds.

From (16),

$$\chi_{\sigma:J}(V) \equiv (-1)^{J-1} \frac{\partial^J \ln G(e^V)}{\partial V_1 \dots \partial V_J} = \int_{-\infty}^{+\infty} \nabla_{1\dots J} R(u - V_1, \dots, u - V_j) du. \tag{19}$$

Then

$$\int_L \chi_{\sigma:J}(V) dV = \int_L \int_{-\infty}^{+\infty} \nabla_{1\dots J} R(u - V_1, \dots, u - V_J) du dV. \tag{20}$$

Make the linear transformation

$$(\zeta_1, \zeta_2, \dots, \zeta_J) = (u, V_2, \dots, V_J) A, \tag{21}$$

with

$$A = \begin{pmatrix} 1 & 1'_{J-1} \\ 1_{J-1} & -I_{J-1} \end{pmatrix}, \tag{22}$$

and note that

$$A^{-1} = J^{-1} 1_J 1'_J + \begin{pmatrix} 0 & 0'_{J-1} \\ 0_{J-1} & -I_{J-1} \end{pmatrix}, \tag{23}$$

and det $A = (-1)^{J-1} J$. Substitute this transformation in (20), noting that the Jacobian is J, and obtain

$$\int_L \chi_{\sigma:J}(V) dV = J^{-1} \int_{-\infty}^{+\infty} \nabla_{1\dots J} R(\zeta_1, \dots, \zeta_J) d\zeta_1 \dots d\zeta_J = J^{-1}. \tag{24}$$

For $j = 1, \dots, J$, the finite mean property for $R(\zeta)$ implies

(25)

$$\begin{aligned} \int_L |V_j| \chi_{\sigma:J}(V) dV &= \int_L \int_{-\infty}^{+\infty} |V_j| \nabla_{1\dots J} R(u - V_1, \dots, u - V_J) du dV \\ &= J^{-1} \int_{-\infty}^{+\infty} |\zeta_j - 1'\zeta/J| \, \nabla_{1\dots J} R(\zeta_1, \dots, \zeta_J) d\zeta_1 \dots d\zeta_J < +\infty. \end{aligned}$$

This completes the demonstration of [G4].

Next, the converse proposition in the theorem will be proved. Assume that $G(e^V)$ is a function satisfying properties [G1]-[G4]. Let $\chi_{\sigma:J}(V)$ be the mixed derivative of $\ln G(e^V)$ from (7), and let η be any continuous density on $\mathbb{R}$ with a finite mean. Consider an ARUM $U_j = V_j + \zeta_j$ in which $\zeta = (\zeta_1, \ldots, \zeta_J)$ is assigned the candidate density

$$\rho(\zeta) = \eta(1'\zeta/J)\chi_{\sigma:J}(1'\zeta/J - \zeta) \geq 0. \tag{26}$$

Transform from $(\zeta_1, \ldots, \zeta_J)$ to $(u, v_2, \ldots, v_J)$, where $u = 1'\zeta/J$ and $v_j = 1'\zeta/J - \zeta_j = u - \zeta_j$ for $j = 1, \ldots, J$. Note that this is just the inverse of the linear transformation (21), and that its Jacobian is $1/J$. Then,

$$v_1 = -\sum_{j=2}^{J} v_j \tag{27}$$

and

$$\int_{-\infty}^{+\infty} \rho(\zeta_1, \ldots, \zeta_J) d\zeta_1 \ldots d\zeta_J = \int_{u=-\infty}^{+\infty} \eta(u) du \int_{v \in L} J\chi_{\sigma:J}(v) dv = 1 \tag{28}$$

by [G4], establishing that ρ is a density. Further, by the finite mean property of η and [G4],

$$\begin{aligned} E_\zeta|\zeta_j| = E_{u,v}|u - v_j| \quad &\leq \quad \int_{u=-\infty}^{+\infty} |u|\eta(u) du \int_{v \in L} J\chi_{\sigma:J}(v) dv \\ &+ \int_{-\infty}^{+\infty} \eta(u) du \int_{v \in L} |v_j| J\chi_{\sigma:J}(v) dv \qquad (29) \\ &< \quad +\infty, \end{aligned}$$

so that this density has a finite mean. Let $R(\zeta)$ denote the CDF for the density ρ. The ARUM $U_j = V_j + \zeta_j$ then has the choice probability

$$
\begin{aligned}
(30)\quad P_{\mathcal{C}}(1|V) &= \Pr(U_1 \geq U_k \text{ for } k = 2, \ldots, J) \\
&= \int_{-\infty}^{+\infty} \nabla_1 R(u - V_1, \ldots, u - V_J) du \\
&= \int_{u=-\infty}^{+\infty} \int_{\zeta_2=-\infty}^{u-V_2} \cdots \int_{\zeta_J=-\infty}^{u-V_J} \rho(u - V_1, \zeta_2, \ldots, \zeta_J) du d\zeta_2 \ldots d\zeta_J \\
[\mathtt{a}] &= \int_{u=-\infty}^{+\infty} \int_{z_2=+\infty}^{V_2} \cdots \int_{z_J=+\infty}^{V_J} \rho(u - V_1, u - z_2, \ldots, u - z_J) du dz_2 \ldots dz_J \\
&= \int_{u=-\infty}^{+\infty} \int_{z_2=+\infty}^{V_2} \cdots \int_{z_J=+\infty}^{V_J} \eta \left(u - (V_1 + \sum_{j=2}^{J} z_j)/J \right) \\
&\quad \chi_{\sigma:J}(V_1 - u, z_2 - u, \ldots, z_J - u) du dz_2 \ldots dz_J \\
[\mathtt{b}] &= \int_{w=-\infty}^{+\infty} \int_{z_2=V_2}^{+\infty} \cdots \int_{z_J=V_J}^{+\infty} \eta(w)(-1)^{J-1} \chi_{\sigma:J}(V_1, z_2, \ldots, z_J) dw dz_2 \ldots dz_J \\
[\mathtt{c}] &= \int_{w=-\infty}^{+\infty} \eta(w) dw \; \chi_1(V_1, V_2, \ldots, V_J) \\
&= \partial \ln G(e^V)/\partial V_1,
\end{aligned}
$$

where in [a] we use the transformation $z_j = u - \zeta_j$ for $j = 2, \ldots, J$; in [b] we use the transformation $w = u - (V_1 + \sum_{j=2}^{J} z_j)/J$; and in [c] we use the condition from [G4] that $\chi_1(z)$ is zero when any component of $(z_2, \ldots, z_J)$ equals $+\infty$. Note that the density η introduces a factor that is common to all utilities and has no effect on the choice probabilities. The result above holds for any permutation of the indices. Thus, $G(e^V)$ is a choice-probability generating function for the ARUM $U_j = V_j + \zeta_j$, completing the proof of the theorem.

■

Since ARUM-consistent choice probabilities are obtained from a gradient of a CPGF, the properties of choice probabilities with this property are easily derived from the properties [G1]-[G4] of a CPGF. Further, it is straightforward to show that these specific properties of choice probabilities are sufficient for ARUMconsistency by integrating back to obtain the associated CPGF. The following corollary gives this result.

Corollary 8 *A choice probability $P_{\mathcal{C}}(j|V)$ on $\mathcal{C} = \{1, \ldots, J\}$ is consistent with a finite-mean TARUM if and only if it is non-negative, sums to one, and satisfies*

[CP1] *For $j \neq k$, $\partial P_{\mathcal{C}}(j|V)/\partial V_k = \partial P_{\mathcal{C}}(k|V)/\partial V_j$, and for every $k = 1, \ldots, J-1$ and permutation σ with $\sigma(J) = j$, so that $\sigma : k$ excludes j,*

$$(31) \qquad (-1)^k \frac{\partial^k P_{\mathcal{C}}(j|V)}{\partial V_{\sigma(1)} \partial V_{\sigma(2)} \ldots \partial V_{\sigma(k)}} \geq 0.$$

[CP2] $P_{\mathcal{C}}(j|V - c) = P_{\mathcal{C}}(j|V)$ *for every scalar c.*

[CP3] *If $\max_{k \neq j} V_k - V_j \to +\infty$, then $P_{\mathcal{C}}(j|V) \to 0$.*

[CP4] *Let $L = \{V \in \mathbb{R}^J | \sum_{j=1}^J V_j = 0\}$. Then*

$$(32) \qquad \int_L (-1)^{J-1} \frac{\partial^{J-1} P_{\mathcal{C}}(j|V)}{\partial V_1 \ldots \partial V_{j-1} \partial V_{j+1} \ldots \partial V_J} dV = J^{-1}$$

and for $j = 1, \ldots, J$,

$$(33) \qquad (-1)^{J-1} \int_L |V_j| \frac{\partial^{J-1} P_{\mathcal{C}}(j|V)}{\partial V_1 \ldots \partial V_{j-1} \partial V_{j+1} \ldots \partial V_J} dV < +\infty.$$

Proof. Given a finite-mean TARUM, let G be an associated CPGF. Properties [G1]-[G4] of G, translated to statements about $P_{\mathcal{C}}(j|V) = \partial G(e^V)/\partial V_j$, become properties [CP1]-[CP4]. Then, by the theorem, [CP1]-[CP4] are necessary for RUM-consistent choice probabilities.

Conversely, given a choice probability $P_{\mathcal{C}}(j|V)$ satisfying [CP1]-[CP4], define a candidate CPGF by

$$(34) \qquad \ln G(e^V) = \sum_{j=1}^{J} \int_{z_j=0}^{V_j} P_{\mathcal{C}}(j|V_1, \ldots, V_{j-1}, z_j, 0, \ldots, 0) dz_j.$$

Then, using the symmetry condition in [CP1]

(35)

$$
\begin{aligned}
\partial \ln G(e^V)/\partial V_k &= P_{\mathcal{C}}(k|V_1,\ldots,V_k,0,\ldots,0) \\
&\quad + \sum_{j=k+1}^{J} \int_{z_j=0}^{V_j} \frac{\partial P_{\mathcal{C}}(j|V_1,\ldots,V_{j-1},z_j,0,\ldots,0)}{\partial V_k} dz_j \\
&= P_{\mathcal{C}}(k|V_1,\ldots,V_k,0,\ldots,0) \\
&\quad + \sum_{j=k+1}^{J} \int_{z_j=0}^{V_j} \frac{\partial P_{\mathcal{C}}(k|V_1,\ldots,V_{j-1},z_j,0,\ldots,0)}{\partial V_j} dz_j \\
&= P_{\mathcal{C}}(k|V_1,\ldots,V_k,0,\ldots,0) \\
&\quad + \sum_{j=k+1}^{J} [P_{\mathcal{C}}(k|V_1,\ldots,V_{j-1},V_j,0,\ldots,0) - P_{\mathcal{C}}(k|V_1,\ldots,V_{j-1},0,0,\ldots,0)] \\
&= P_{\mathcal{C}}(k|V_1,\ldots,V_k,\ldots,V_J).
\end{aligned}
$$

Then G generates the given choice probability. The proof is completed by showing that G defined by (34) satisfies [G1]-[G4]. The argument above establishes that $\chi_j(V) \equiv P_{\mathcal{C}}(j|V) \geq 0$. Then [CP1] implies [G1]. From (35),

$$
\sum_{k=1}^{J} \frac{\partial \ln G(e^V)}{\partial V_k} = \sum_{k=1}^{K} e^{V_k} \nabla_k G(e^V)/G(e^V) = \sum_{k=1}^{J} P_{\mathcal{C}}(k|V) = 1,
$$

implying that

$$
\sum_{k=1}^{J} e^{V_k} \frac{\partial G(e^V)}{\partial V_k} = G(e^V),
$$

and hence by Euler's theorem that G satisfies [G2]. Conditions [G3] and [G4] follow from [CP3] and [CP4] by noting that $\chi_j(V) \equiv P_{\mathcal{C}}(j|V)$. Then, the candidate CPGF G satisfies [G1]-[G4] and generates the given choice probability. ■

Prior literature: McFadden (1981) proves a theorem, attributed to Williams (1977) and Daly and Zachary (1978), that utilizes an economic argument to establish necessary and sufficient conditions under which RUM models with additive income terms in their utilities can be characterized equivalently in terms of a social surplus function whose gradient gives the choice probabilities. Theorem 7 and Corollary 8 use related arguments with location parameters V in place of income, and obtain similar but more complete results, without restrictive economic conditions on utility. Papers that deal with the relationship between characterization of choice models and integrability conditions in economic demand theory are Bates (2003), Ibáñez (2007), Ibáñez and Batley (2008), Koning and Ridder (2003), and McFadden (2005). Papers that deal with choice-probability generating functions for nested logit and what econometricians term generalized extreme value models are McFadden (1978), Daly and Bierlaire (2006) and Bierlaire, Bolduc and McFadden (2008). A paper that considers mixing of choice models is McFadden and Train (2000).

2.3 Mixing of CPGF

Suppose a TARUM $\varepsilon_j = r_j(U_j - V_j, \alpha)$ in which $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_J)$ has CDF $F(\varepsilon)$ and α is a latent mixing vector, independent of ε, with density $K(\alpha)$. The usual interpretation of α is that it indexes the tastes of heterogeneous consumers. Let $G(e^V, \alpha)$ denote the CPGF, which is measurable as a function of α, and $P_{\mathcal{C}}(j|V, \alpha) = \partial G(e^V, \alpha)/\partial V_j$ the choice probability associated with the TARUM given α. Then there is an observationally equivalent mixed ARUM $U_j = V_j + \zeta_j$ with ζ having a CDF $R(\zeta|\alpha)$ given α, or an unconditional CDF $R(\zeta) = \int R(\zeta|\alpha)K(\,\mathrm{d}\alpha)$. The associated choice probabilities are the corresponding mixtures $P_{\mathcal{C}}(j|V) = \int P_{\mathcal{C}}(j|V, \alpha)K(\,\mathrm{d}\alpha)$; these are obviously RUM-consistent. A simple *sum log* formula gives the CPGF for this mixed model:

$$\ln G(e^V) = \int \ln G(e^V, \alpha) K(\,\mathrm{d}\alpha). \tag{36}$$

It has a gradient that gives the mixed choice probabilities. It is easy to verify that a mixture of logs of functions $G(e^V, \alpha)$ that satisfy [G1]-[G4] is the log of a function that again satisfies these conditions, and hence is a CPGF for the mixed model. This result can be used together with the results of McFadden and Train (2000) that every regular RUM can be approximated as closely as one pleases by some mixture of logit models (or other base models such as independent probit); the corresponding sumlog mixture of the CPGF for the base model will give the RUM-consistent CPGF corresponding to the approximating probabilities.

2.4 Economic variables and economic rationality

Suppose consumers are rational economic decision-makers that seek maximum utility through discrete and continuous consumption choices. Let $j \in \mathcal{C}$ denote the set of discrete alternatives, and t_j denote the cost of alternative j. Let $x \in X$ denote a continuous consumption vector, and let p denote a corresponding price vector. Let y denote income. Then, the consumer with a utility function $u(x, j)$ will solve $\max_{j \in \mathcal{C}} \max_{x \in X} u(x, j)$ subject to the budget constraint $p'x \leq y - t_j$, where the right-hand-side is the consumer s discretionary income after paying for discrete alternative j. The solution to the inner maximization, $U_j(p, y - t_j) = \max_{x \in X}\{u(x, j)|p'x \leq y - t_j\}$, is termed *indirect utility* conditioned on j. The indirect utility function is homogeneous of degree zero and quasi-convex[1] in $(p, y - t_j)$, increasing in $y - t_j$, and non-increasing in p, and any function with these properties induces a utility function $u(x, j) = \inf_p U_j(p, p'x)$

[1] A function $U_j(p, y)$ is quasi-convex if its lower contour set $\{(p, y)|U_j(p, y) \leq u\}$ is convex for each u, and quasi-concave if its negative is quasi-convex.

for which U_j is the associated indirect utility function; cf. Diewert (1974), McFadden (1974), Varian (1993, Sect. 7.2). The indirect utility U_j may vary over a population due to heterogeneity in tastes, and in addition to its dependence on $(p, y - t_j)$, it may depend on other variables X_j that include attributes of the discrete alternative and factors that influence tastes, such as age and family size. The economic rationality requirement of consumer sovereignty, under which tastes are determined and invariant with respect to the economic budget the consumer faces, requires that any latent taste factors in U_j have distributions that are independent of the budget $(p, y - t_j)$. In a well-specified economically rational TARUM, this corresponds to a specification $\varepsilon_j = r_j(U_j - V_j, p, y - t_j, X_j)$ in which the vector of random taste factors $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_J)$ has CDF $F(\varepsilon)$ that is independent of $(p, y - t_j)$. For this TARUM to define an indirect utility function satisfying the homogeneity, quasi-convexity, and monotonicity requirements of economic rationality, $r_j(U_j - V_j, p, y - t_j, X_j)$ must be increasing in $U_j - V_j$, homogeneous of degree zero and quasi-concave in $(p, y - t_j)$, non-decreasing in p, and decreasing in y.[2]

Our specification of economically rational TARUM in which the location pa-

[2] If $p^0 \leq p^1$ and $y^0 - t_j^0 \geq y^1 - t_j^1$, then every continuous commodity vector available at the budget $(p^1, y^1 - t_j^1)$ is also available at $(p^0, y^0 - t_j^0)$. Then, $(p^0, y^0 - t_j^0)$ must yield at least as high a utility, implying $r_j(U_j - V_j, p^0, y^0 - t_j^0, X_j, \alpha) \leq r_j(U_j - V_j, p^1, y^1 - t_j^1, X_j, \alpha)$ for each U_j. If $y^0 - t_j^0 > y^1 - t_j^1$ and the consumer can always increase utility by exhausting disposable income, then $(p^0, y^0 - t_j^0)$ must yield higher utility, implying that r_j is strictly decreasing in income. If $(p^0, y^0 - t_j^0)$ and $(p^1, y^1 - t_j^1)$ are two budgets, and $(p^\theta, y^\theta - t_j^\theta) = \theta(p^0, y^0 - t_j^0) + (1 - \theta)(p^1, y^1 - t_j^1)$ for some $0 < \theta < 1$, then the continuous commodity vector that maximizes utility at the budget $(p^\theta, y^\theta - t_j^\theta)$ must be affordable at either the budget $(p^0, y^0 - t_j^0)$ or the budget $(p^1, y^1 - t_j^1)$, implying that the utility at one of these two budgets is at least as high as the utility at the budget $(p^\theta, y^\theta - t_j^\theta)$. Then, for each U_j, $r_j(U_j - V_j, p^\theta, y^\theta - t_j^\theta, X_j, \alpha)$ must be at least as large as the minimum of $r_j(U_j - V_j, p^0, y^0 - t_j^0, X_j, \alpha)$ and $r_j(U_j - V_j, p^1, y^1 - t_j^1, X_j, \alpha)$. This shows that r_j is quasi-concave.

rameters V_j are independent of economic variables appears incompatible with the common practice of introducing dependence on income and prices through a "systematic" utility component V_j. However, the TARUM transformation r_j may include an additional quasi-concave function of discretionary income and prices that is additive to V_j and plays the conventional role. For example, the transformation

$$(37) \quad r_j(U_j - V_j, p, y - t_j, X_j, \alpha) = \\ \nu(U_j - V_j - [(y - t_j)/a(p)]^{1-\eta}/(1-\eta) + b_j(p, X_j, \alpha)/a(p))$$

with ν an increasing transformation, $a(p)$ and $b_j(p, X_j, \alpha)$ concave linear homogeneous nondecreasing functions of p, $\eta \neq 1$ a positive parameter, and α a latent stochastic taste factor, defines a useful class of indirect utility functions, called the generalized Gorman preference field (Gorman, 1953), that allow discrete responses to depend on income level and on the prices of continuous goods.

3 RUM with multivariate extreme value distributed utility

3.1 Multivariate Extreme Value distributions

The logit, nested logit, and cross-nested logit choice probabilities that have been important for applications are consistent with finite-mean RUM that have multivariate extreme value (MEV) distributions; see McFadden (1974) and McFadden (1978). In this section, we summarize selected results from the statistical theory of MEV distributions that allow us to relate MEV-RUM families to the general characterization of finite-mean RUM-consistent CPGF and choice probabilities given in Theorem 7 and its corollary.

Definition 9 *If $F(\varepsilon)$ is an absolutely continuous CDF on $\mathbb{R}^J$ with univariate* *marginals* $F_{(j)}(\varepsilon_j)$*, and copula*[3]

$$C(u_1, \ldots, u_J) = F(F_{(1)}^{-1}(u_1), \ldots, F_{(J)}^{-1}(u_J))$$

on $[0,1]^J$*, define the function* $A : [0,+\infty]^J \to [0,+\infty]$ *by*

(38) $$A(y_1, \ldots, y_J) \equiv -\ln C(e^{-y_1}, \ldots, e^{-y_J}),$$

and term A *the* exponent *of* C *or equivalently of* F*; then*

(39) $$C(u_1, \ldots, u_J) = e^{-A(-\ln u_1, \ldots, -\ln u_J)}$$

and

(40) $$F(\varepsilon) = e^{-A\left(-\ln F_{(1)}(\varepsilon_1), \ldots, -\ln F_{(J)}(\varepsilon_J)\right)}.$$

Remark The exponent of a multivariate CDF is not in general an associated CPGF, but we will later establish an equivalence that when F is a multivariate extreme value distribution then A is in fact a CPGF, satisfying the conditions [G1]-[G4] and generating the choice probabilities for a TARUM based on F. The exponent of a multivariate distribution is determined solely by its copula, independent of its univariate marginals. Hence, the mapping from multivariate distributions to exponents is many-to-one, with the inverse image of a given exponent being the family of all multivariate distributions with the same copula and various univariate marginals.

The following mathematical results will be used to obtain the features of the exponent of a CDF, and its relation to CPGF.

[3] A copula is a CDF with uniform [0,1] univariate marginals. With the transformation given, every absolutely continuous multivariate CDF can be written $F(\varepsilon_1, \ldots, \varepsilon_J) = C(F_{(1)}(\varepsilon_1), \ldots, F_{(J)}(\varepsilon_J))$, and hence is characterized by its univariate marginals and its copula.

Definition 10 *A real-valued function ψ of a scalar t is completely monotonic if, for $k = 1, 2, \dots,$*

$$(41)\qquad (-1)^{k-1}\partial^k\psi(t)/\partial t^k \geq 0.$$

Examples of completely monotonic functions (see Miller and Samko, 2001) are:

- $\ln(\alpha + \beta t)$*, with* $\alpha + \beta t > 0$ *and* $\beta > 0$*;*
- t^α*, with* $t \geq 0$ *and* $0 \leq \alpha \leq 1$*;*
- $-e^{-\alpha t}$*, with* $\alpha \geq 0$*;*
- $-1/t$*, with* $t < 0$.

Lemma 11 *If a real-valued function g on $[0, +\infty)^J$ satisfies ASP from Definition 5, and if ψ is completely monotonic on a domain that includes the range of g, then $\psi(g)$ satisfies ASP.*

Proof. A mixed derivative of $\psi(g)$ can be written in terms of mixed derivatives of g as

$$(42)\qquad \nabla_{\sigma:k}\psi(g(y)) = \sum_{i=1}^{k}\psi^{(i)}(g(y)) \sum_{(a_1,\dots,a_i)\in A^i_{\sigma_k}} \nabla_{a_1}g(y)\dots\nabla_{a_i}g(y),$$

where $\psi^{(i)}(t) = \partial^i\psi(t)/\partial t^i$ and $A^i_{\sigma:k}$ is the set of all partitions of $(\sigma(1),\dots,\sigma(k))$ into i non-empty subvectors. This formula can be verified by induction. The ASP follows by induction, since differentiation of a term in (42) leads to new terms, all having the opposite sign of the original term. ■

Corollary 12 *If $A : [0, +\infty]^J \to [0, +\infty]$ satisfies ASP, then*

- $\ln A$ *satisfies ASP;*
- αA *satisfies ASP if* $\alpha > 0$*;*

- *A^α satisfies ASP if $0 < \alpha \leq 1$;*

Proof. From Lemma 11 and the fact that $\psi(t) = \ln(t)$, $\psi(t) = \alpha t$ and $\psi(t) = t^\alpha$ are completely monotonic for $t > 0$. ■

Corollary 13 *If $A_1 : [0, +\infty]^J \to [0, +\infty]$ and $A_2 : [0, +\infty]^J \to [0, +\infty]$ satisfies ASP, then*

- *$\alpha A_1 + \beta A_2$ satisfies ASP if $\alpha, \beta \geq 0$;*
- *$A_1 A_2$ satisfies ASP;*

Proof. The results comes from Lemma 11 and the properties of completely monotonic (c.m.) functions that a linear combination of c.m. functions with non-negative multipliers is also c.m., and that the product of two c.m. functions is also c.m. ■

Lemma 14 *If $A : [0, +\infty]^J \to [0, +\infty]$ is the exponent of an absolutely continuous CDF $F(\varepsilon)$, then A satisfies [G3], and satisfies ASP and therefore [G1].*

Proof. Consider the construction $A(y) = -\ln C(e^{-y})$. Computation shows that $-C(e^{-y})$ satisfies ASP, that $\psi_1(t) = -1/t$ is completely monotonic for $t < 0$, and that $\psi_2(t) = \ln t$ is completely monotonic for $t > 0$. Then by Lemma 11, the composition $A(y) = \psi_2(\psi_1(-C(e^{-y}))$ satisfies ASP. Then, Corollary 12 implies that $\ln A(y)$ satisfies ASP, and hence [G1]. The boundary conditions $C(1) = 1$ and $C(u_1, \ldots, u_{j-1}, 0, u_{j+1}, \ldots, u_J) = 0$ for each j imply that A satisfies [G3]. ■

Remark The function $G(y_1, y_2) = \sqrt{y_1 + 2y_2}\sqrt{2y_1 + y_2}$ satisfies [G1] (i.e., ASP for $\ln G$), but not ASP for G. Then, ASP for G is strictly stronger than [G1].

Definition 15 *Let $\varepsilon \in \mathbb{R}^J$ be a random vector with CDF $F(\varepsilon)$ and let $m_n \equiv \bigvee_{k\leq n} \varepsilon_k = (\bigvee_{k\leq n} \varepsilon_{k1}, \ldots, \bigvee_{k\leq n} \varepsilon_{kJ})$ be the component-wise maximum of independent draws $\varepsilon_1, \ldots, \varepsilon_n$ from F. Then ε is* location/scale max-stable *if there are location and scale parameter vectors, $\mu(n)$ and $\sigma(n)$ respectively, such that the CDF of m_n is $F((m_{n1} - \mu_1(n))/\sigma_1(n), \ldots, (m_{nJ} - \mu_J(n))/\sigma_J(n))$.*

Definition 16 *The generalized extreme value (GEV) distribution has CDF*

$$H(\varepsilon; \xi, \mu, \sigma) = \exp\left(-\left[1 + \xi\frac{\varepsilon - \mu}{\sigma}\right]_+^{-\frac{1}{\xi}}\right),$$

where $\sigma > 0$, $[x]_+ = \max(x, 0)$ and $H(\varepsilon; 0, \mu, \sigma) = \lim_{\xi\to 0} H(\varepsilon; \xi, \mu, \sigma) =$ $\exp\left(-\exp\left(-\frac{\varepsilon-\mu}{\sigma}\right)\right)$. H is max-stable since, for $\lambda > 0$, $\mu(\lambda) = (1 - \lambda^\xi)\frac{\xi\mu-\sigma}{\xi}$ and $\sigma(\lambda) = \lambda^\xi$, with limits $\mu(\lambda) = \sigma \ln \lambda$ and $\sigma(\lambda) = 1$ as $\xi \to 0$, satisfy $H(\varepsilon; \xi, \mu, \sigma)^\lambda = H\left(\frac{\varepsilon-\mu(\lambda)}{\sigma(\lambda)}; \xi, \mu, \sigma\right)$.

Remark By the Fisher - Tippet - Gnedenko theorem, any max-stable univariate distribution belongs to the GEV family (Joe, 1997). The distribution $H(\varepsilon; 0, \mu, \sigma)$ is known as the type 1 extreme value (EV1) or Gumbel distribution, which has mean $\mu + \sigma\gamma$, where γ is Euler's constant. When $\xi > 0$, $H(\varepsilon; \xi, \mu, \sigma)$ is called type 2 extreme value or Fréchet. When $\xi < 0$, $H(\varepsilon; \xi, \mu, \sigma)$ is called type 3 extreme value or reverse Weibull.

Definition 17 *Let $\varepsilon \in \mathbb{R}^J$ be a random vector with CDF $F(\varepsilon)$. Then ε is termed* multivariate extreme value *(MEV) if it is location/scale max-stable. The univariate marginals of a MEV distribution must be in the GEV family just defined.*[4]

[4]We follow the statistics literature in naming the univariate family $H(\varepsilon; \xi, \mu, \sigma)$ the *generalized extreme value* (GEV) distribution; the econometrics literature often uses the term GEV for MEV distributions. Characterization of univariate max-stable distributions dates to Fisher and Tippett (1928), Gnedenko (1943), and Gumbel (1958); see also Johnson et al. (1995). Joe (1997) and Coles (2001) give detailed treatments of multivariate max-stable distributions. When transformations other than location and scale are allowed, the family of max-stable distributions has additional members; see Sreehari (2009).

A max-stable CDF $F(\varepsilon_1, \ldots, \varepsilon_J)$ must then satisfy

$$
\begin{aligned}
&F(\varepsilon_1, \ldots, \varepsilon_J)^\lambda \equiv \\
&C(H(\varepsilon_1; \xi_1, \mu_1, \sigma_1), \ldots, H(\varepsilon_J; \xi_J, \mu_J, \sigma_J))^\lambda \equiv \\
&C\left(H\left(\tfrac{\varepsilon_1 - \mu_1(\lambda)}{\sigma_1(\lambda)}; \xi_1, \mu_1, \sigma_1\right), \ldots, H(\tfrac{\varepsilon_J - \mu_J(\lambda)}{\sigma_J(\lambda)}; \xi_J, \mu_J, \sigma_J)\right) \equiv \\
&C(H(\varepsilon_1; \xi_1, \mu_1, \sigma_1)^\lambda, \ldots, H(\varepsilon_J; \xi_J, \mu_J, \sigma_J)^\lambda).
\end{aligned} \tag{43}
$$

This demonstrates the following result:

Lemma 18 *(Joe, 1997) $F(\varepsilon_1, \ldots, \varepsilon_J)$ is a MEV distribution if and only if its copula C satisfies the homogeneity condition*

$$C(u_1, \ldots, u_J)^\lambda = C(u_1^\lambda, \ldots, u_J^\lambda) \tag{44}$$

for $u \in [0,1]^J$ and $\lambda > 0$.

Remark If $C(u_1, \ldots, u_J)$ is the copula of any multivariate CDF $F(\varepsilon_1, \ldots, \varepsilon_J)$, then other multivariate distributions with the same copula can be obtained by substituting different univariate marginal distributions into the copula. In particular, if C is the copula of a MEV distribution $F(\varepsilon_1, \ldots, \varepsilon_J)$, so that C has the homogeneity property in Lemma 18, then

$$F^*(\varepsilon_1, \ldots, \varepsilon_J) = C\left(e^{-e^{-\varepsilon_1}}, \ldots, e^{-e^{-\varepsilon_J}}\right)$$

is a MEV distribution with the same copula and EV1 univariate marginals with location parameters zero and scale parameters 1. We will term these *MEV1 distributions*.

Combined with the definition (38) of the exponent of a distribution, Lemma 18 gives the following result:

Lemma 19 *If $A : [0, +\infty]^J \to [0, +\infty]$ is the exponent of an absolutely continuous CDF $F(\varepsilon)$, then A satisfies [G2] if and only if F is max-stable.*

Proof. $A(y) = -\ln C(e^{-y})$ implies $A(\lambda y) = -\ln C\left((e^{-y})^{\lambda}\right) = -\ln C(e^{-y})^{\lambda} = \lambda A(y)$. ■

Not all multivariate max-stable distributions have the property that vectors of cross-alternative maxima are max-stable. For example, $F(\varepsilon_1, \varepsilon_2) = H(\varepsilon_1; 0, 0, 1)H(\varepsilon_2; 0, 0, 2)$ implies that $M = \max(\varepsilon_1, \varepsilon_2)$ has the CDF $\Pr(M \leq m) = \exp(-e^{-m} - e^{-m/2})$, which is not a max-stable distribution. The following result gives the condition under which max-stability is preserved under cross-alternative maximization; we term this *cross-alternative max-stability* (CAMS).

Lemma 20 *Suppose $F(\varepsilon_1, \ldots, \varepsilon_J)$ is max-stable. For every j, k, $\max(\varepsilon_j+V_j, \varepsilon_k+V_k)$ is max-stable for all location-shift parameters V_j, V_k if and only if the univariate marginals of F are EV1 with a common scale parameter σ.*

Proof. If the univariate marginals of F are all EV1 with common scale parameter σ, Lemma 19 gives

$$\Pr\left(\max_{j\leq J}(\varepsilon_j + V_j) \leq m\right) = e^{-A\left(e^{-(m-\mu_1-V_1)/\sigma},\ldots,e^{-(m-\mu_J-V_J)/\sigma}\right)} = e^{-e^{-m/\sigma}A\left(e^{(\mu_1+V_1)/\sigma},\ldots,e^{(\mu_J+V_J)/\sigma}\right)}, \tag{45}$$

and the cross-alternative maxima are EV1 distributed with mean

$$\sigma \ln A\left(e^{(\mu_1+V_1)/\sigma}, \ldots, e^{(\mu_J+V_J)/\sigma}\right) + \sigma\gamma.$$

Conversely, if $\max(\varepsilon_j + V_j, \varepsilon_k + V_k)$ is max-stable, then for some max-stable univariate distribution $H(\varepsilon; \xi, \mu, \sigma)$,

$$\Pr\left(\max(\varepsilon_j + V_j, \varepsilon_k + V_k) \leq m\right) = e^{-A(+\infty,\ldots,\infty,-\ln H(m-V_j;\xi_j,\mu_j,\sigma_j),+\infty,\ldots,+\infty,-\ln H(m-V_k;\xi_k,\mu_k,\sigma_k),+\infty,\ldots,+\infty)} \equiv H(m;\xi;\mu;\sigma). \tag{46}$$

The homogeneity of A implies that

$$0 \equiv A\left(+\infty, \ldots, \infty, L_j, +\infty, \ldots, +\infty, L_k, +\infty, \ldots, +\infty\right),$$

with

$$L_i = \ln \frac{H(m; \xi; \mu; \sigma)}{H(m - V_i; \xi_i, \mu_i, \sigma_i)}, \quad i = j, k.$$

The first finite argument of A is independent of m for more than one value of V_j if and only if $\xi = \xi_j = 0$ and $\sigma = \sigma_j$. Since A is increasing in each of its finite arguments for some range of location shifts, both arguments must be independent of m, requiring that $\xi = \xi_j = \xi_k = 0$ and $\sigma_k = \sigma_j$. ■

Remark The following functions are examples of MEV exponents.

- $A\left(y\right) = \left(\sum_{j \in \mathcal{C}} y_j^{\delta}\right)^{\frac{1}{\delta}}$ for $\delta \geq 1$.
- $$A\left(y\right) = \sum_{T \subseteq \mathcal{C}} (-1)^{|T|+1} \left[\sum_{i \in T} y_i^{-\delta}\right]^{-\frac{1}{\delta}}, 0 \leq \delta \leq \infty.$$

More examples exist, some are indicated in Joe (1997).

3.2 CPGF and CAMS MEV

We will show that a random utility model characterized by an ARUM $U_j = V_j + \zeta_j$ with a CDF $R(\zeta_1, \ldots, \zeta_J)$, or an observationally equivalent TARUM, has a CPGF given by the exponent of R if and only if R is a CAMS MEV. If R is MEV but not CAMS, then its exponent is a CPGF for another ARUM, not observationally equivalent, that is CAMS with the same copula. We state the conditions on a CPGF, or on a choice probability, that are necessary and sufficient for them to be associated with a cross-alternative max-stable ARUM. We tie these results to the analysis by McFadden (1978) and Smith (1984) of choice probabilities derivable from "generalized extreme value" models.

The following two assumptions are stronger than the parallel assumptions [G1] and [CP1], and are used to establish that a candidate CPGF $G : [0,+\infty]^J \to [0,+\infty]$, or a choice probability $P_{\mathcal{C}}(j|V)$, is consistent with a CAMS ARUM:

[G1*] (Strong ASP) $G(y)$ satisfies ASP.

[CP1*] For $j \neq k$, $\partial P_{\mathcal{C}}(j|V)/\partial V_k = \partial P_{\mathcal{C}}(k|V)/\partial V_j$. For $k = 1,\ldots,J-1$ and permutation σ satisfying $\sigma(J) = j$, the test functions $T_{\sigma:k}(V)$, defined recursively by

- $T_{\sigma(1)}(V) = P_{\mathcal{C}}(j|V)$ and,
- for $k > 1$, $T_{\sigma:k+1}(V) = T_{\sigma:k}(V)T_{\sigma(k+1)}(V) + \partial T_{\sigma:k}(V)/\partial V_{k+1}$,

satisfy

$$(-1)^{k-1}T_{\sigma:k}(V) \geq 0.$$

Theorem 21 *Suppose $A : [0,+\infty]^J \to [0,+\infty]$ is the exponent of a multivariate CDF $R(\zeta_1,\ldots,\zeta_J)$. Then, A satisfies [G3], and [G1*] and hence [G1]. Second, A satisfies [G2] if and only if $R(\zeta_1,\ldots,\zeta_J)$ is MEV. Third, A is a CPGF for the ARUM $U_j = V_j + \zeta_j$ with a MEV CDF $R(\zeta_1,\ldots,\zeta_J)$ if and only if $R(\zeta_1,\ldots,\zeta_J)$ is CAMS, but any A satisfying [G1*], [G2], and [G3] is a CPGF for an ARUM $U_j = V_j + \nu_j$ with the CAMS MEV $\tilde{R}(\nu_1,\ldots,\nu_J) = \exp(-A(\exp(-\nu_1),\ldots,\exp(-\nu_J)))$. In these results, [G4] is implied by [G1*], [G2], and [G3]. In general, the ARUM $U_j = V_j + \zeta_j$ and $U_j = V_j + \nu_j$ are not observationally equivalent.*

Proof. Lemma 14 establishes that the exponent of R satisfies [G3], [G1*], and hence [G1]. Lemma 19 establishes that A satisfies [G2] if and only if R is MEV. Lemma 20 establishes that A is a CPGF for the ARUM $U_j = V_j + \zeta_j$ with a MEV CDF $R(\zeta_1,\ldots,\zeta_J)$ if and only if $R(\zeta_1,\ldots,\zeta_J)$ is CAMS. McFadden (1978) proves that if A satisfies [G1*], [G2], and [G3], then it is a CPGF for the ARUM $U_j = V_j + \nu_j$ with the CAMS MEV distribution $R(\nu_1, \ldots, \nu_J) = \exp(-A(\exp(-\nu_1), \ldots, \exp(-\nu_J)))$. Since the CAMS MEV distribution has finite means, [G4] is implied by the first part of Theorem 7. ■

Remark Suppose R is a CDF corresponding to a RUM and that R is MEV. By the remarks in Section 2.1, this has a representation that is embedded in an observationally equivalent TARUM with EV1 marginals. This TARUM is MEV and hence it is also CAMS MEV.

If one requires cross-alternative max-stability in the presence of scale shifts rather than location shifts, one finds that Weibull univariate marginals with a common shape parameter $\xi < 0$ and location parameter $\mu = 0$ are CAMS for the family of scale shifts. This does not create a new family of ARUM and CPGF, as one is obtained from the other by a log transformation of utility. This is utilized in section 4 on survival models.

Lemma 22 [CP1*] *for the choice probability $P_{\mathcal{C}}(j|V)$ implies [G1*] for the function*

(47) $$\ln G(e^V) = \sum_{j=1}^{J} \int_{z_j=0}^{V_j} P_{\mathcal{C}}(j|V_1, \ldots, V_{j-1}, z_j, 0, ..., 0)\, \mathrm{d}z_j,$$

and implies [CP1].

Proof. Using the symmetry condition in [CP1*], (47) implies

$$
\begin{aligned}
\frac{\partial \ln G(e^V)}{\partial V_k} &= P_{\mathcal{C}}(k|V_1,\ldots,V_k,0,\ldots,0) \\
&\qquad + \sum_{j=k+1}^{J} \int_{z_j=0}^{V_j} \frac{\partial P_{\mathcal{C}}(j|V_1,\ldots,V_{j-1},z_j,0,\ldots,0)}{\partial V_k}\,\mathrm{d}z_j \\
&= P_{\mathcal{C}}(k|V_1,\ldots,V_k,0,\ldots,0) \\
&\qquad + \sum_{j=k+1}^{J} \int_{z_j=0}^{V_j} \frac{\partial P_{\mathcal{C}}(k|V_1,\ldots,V_{j-1},z_j,0,\ldots,0)}{\partial V_j}\,\mathrm{d}z_j \\
&= P_{\mathcal{C}}(k|V_1,\ldots,V_k,0,\ldots,0) \\
&\quad + \sum_{j=k+1}^{J} \left[P_{\mathcal{C}}(k|V_1,\ldots,V_{j-1},V_j,0,\ldots,0) - P_{\mathcal{C}}(k|V_1,\ldots,V_{j-1},0,0,\ldots,0)\right] \\
&= P_{\mathcal{C}}(k|V_1,\ldots,V_J).
\end{aligned}
$$

Apply the derivative formula (42) to $g(y) = \ln G(y)$ and $\psi(t) = e^t$ to obtain

$$
\frac{\nabla_{\sigma:k} G(y)}{G(y)} = \sum_{i=1}^{k} \sum_{(a_1,\ldots,a_i)\in A^i_{\sigma:k}} \nabla_{a_1} \ln G(y) \cdots \nabla_{a_i} \ln G(y). \tag{48}
$$

Differentiate this expression with respect to $y_{\sigma(k+1)}$ to obtain the recursion

$$
\frac{\nabla_{\sigma:k+1} G(y)}{G(y)} = \frac{\nabla_{\sigma:k} G(y)}{G(y)} \frac{\nabla_{\sigma(k+1)} G(y)}{G(y)} + \frac{\partial \nabla_{\sigma:k} G(y)/\partial y_{\sigma(k+1)}}{G(y)}. \tag{49}
$$

Substitute $y = e^V$ in (49) for each k, multiply this expression by $\exp(V_{\sigma(1)} + \cdots + V_{\sigma(k)})$, and name it $T_{\sigma:k}(V)$.Then (49) becomes the recursion $T_{\sigma:k+1}(V) = T_{\sigma:k}(V)T_{\sigma(k+1)}(V) + \partial T_{\sigma:k}(V)/\partial V_{k+1}$ in [CP1*]. The construction of G gives $T_{\sigma(1)}(V) = P_{\mathcal{C}}(\sigma(1)|V)$. Then, the condition $(-1)^{k-1}T_{\sigma:k}(V) \geq 0$ in [CP1*] and the equality

$$
\exp(V_{\sigma(1)} + \cdots + V_{\sigma(k)})\nabla_{\sigma:k} G(e^V)/G(e^V) = T_{\sigma:k}(V)
$$

implies [G1*]. By Lemma (12), [G1*] implies [G1], and [G1] and the condition $\partial \ln G(e^V)/\partial V_j = P_{\mathcal{C}}(j|V)$ establish [CP1]. ∎

Lemma 23 *Smith (1984) If a choice probability $P_{\mathcal{C}}(j|V)$ satisfies [CP1*], [CP2], and [CP3], then it is consistent with an ARUM $U_j = V_j + \nu_j$ with the CAMS MEV*

$$R(\nu_1, \ldots, \nu_J) = \exp\left(-G\left(e^{-\nu_1}, \ldots, e^{-\nu_J}\right)\right),$$

with G given by (47).

Prior literature: Building on Strauss (1979) and Robertson and Strauss (1981), Lindberg et al. (1995) discuss ARUM that have the property of invariance of achieved utility (IAU). This property states that the distribution of utility conditional on choosing alternative j is independent of j. They show that this class of models coincides with the class of models where the CDF of utility has the form $F(x) = \phi(G(\exp(-x)))$, where G satisfies [G2] (homogeneity). See also de Palma and Kilani (2007)

Remark It is straightforward to establish that the CPGF corresponding to such a distribution is in fact G, that G satisfies strong ASP, and that any ARUM with the IAU property is observationally equivalent to a CAMS ARUM. Admissible functions ϕ may be constructed as Laplace transforms of random variables, see Joe (1997, pp. 204). In this case, F may be viewed as a power mixture of a MEV distribution or equivalently as the distribution obtained from adding the same random variable to all utilities in a MEV distribution. This has no effect on choice probabilities. In the converse direction, one may use that a function ψ on $[0, \infty[$ is the Laplace transform of a random variable if and only if ψ is completely monotonic and $\psi(0) = 0$ (Nelsen, 2006, Lemma 4.6.5). The copula corresponding to $F(x) = \phi(G(\exp(-x)))$ is $\phi(G(\phi^{-1}(u)))$. The case $G(y) = y_1 + ... + y_J$ corresponds to the Archimedean copula (Nelsen, 2006). In the case of a general CPGF G, the corresponding copula may thus be called a *generalized Archimedean copula*.

3.3 An approximation result

This section establishes a constructive way to approximate any ARUM by a certain type of MEV ARUM. It is known that any RUM may be approximated arbitrarily well by a MEV model. This section will show that the choice probabilities of any ARUM, $U_i = V_i + \varepsilon_i$, may be approximated arbitrarily well by the choice probabilities of a MEV model from the nested logit family, for V in a compact set. In fact, the approximating type of model is a cross-nested logit model (Vovsha, 1997, Bierlaire, 2006). The proof of this combines a theorem by Dagsvik (1995) and its proof with the following Lemma.

Lemma 24 *Let μ be a finite measure on $\mathbb{R}^J$. Let $\{g_k\}$ by a finite set of continuous and $\mu-$integrable functions from $\mathbb{R}^J$ to $\mathbb{R}$. Let $K \subset \mathbb{R}^J$ be a compact set. Then for all $\delta > 0$ there exist $\left\{w_n^N\right\}_{n=1}^N, \left\{x_n^N\right\}_{n=1}^N$ such that for every k and for every $V \in K$*

$$\left| \int g_k \left(x - V\right) \mu \left(dx\right) - \sum_{n=1}^N w_n^N g_k \left(x_n^N - V\right) \right| < \delta.$$

Proof. Consider $m \in \mathbb{N}$ and divide $\mathbb{R}^J$ into cubes C_n^m with sides of length 2^{-m}. Choose a point x_n^m in each cube and let $w_n^m = \mu\left(C_n^m\right)$. This construction does not depend on g_k and V.

Define $g_k^m\left(x\right) = g_k\left(x_n^m\right)$ when $x \in C_n^m$ and note that g_k is the pointwise limit of g_k^m since g_k is continuous. Let $g = sup\{g_k(x)|k, x \in K\} < \infty$ and note that $\int g\mu(dx) < \infty$. Then by the Lebesgue dominated convergence theorem

$$\begin{aligned}
\infty &> \int g_k \left(x - V\right) \mu \left(dx\right) \\
&= \lim_{m\to\infty} \int g_k^m \left(x - V\right) \mu \left(dx\right) \\
&= \lim_{m\to\infty} \sum_n w_n^m g_k \left(x_n^m\right) \\
&= \lim_{m\to\infty} \lim_{N\to\infty} \sum_n^N w_n^m g_k \left(x_n^m\right)
\end{aligned}$$

This shows that $\sum_{n=1}^{N} w_n^N g_k\left(x_n^N - V\right)$ tends to $\int g_k\left(x - V\right)\mu\left(dx\right)$ for every k and for every V. Moreover, $\int g_k\left(x - V\right)\mu\left(dx\right)$ is continuous in V. Then convergence is uniform over the compact set $\{1, .., k\} \times K$, which proves the Lemma.

■

We shall now review Theorem 1 in Dagsvik (1995). Dagsvik's theorem applies to general random utility. Here we specialize the result to the additive case $U_i = V_i + \varepsilon_i$, where ε follows some general multivariate distribution with CDF F. This means that the mapping from V to the CDF of U is continuous. The corresponding choice probabilities are $P^F\left(j|V\right)$. Dagsvik assumes that there exists a small $\bar{a} > 0$ such that $0 < Ee^{a\varepsilon_j} < \infty$ for any j and any $0 < a < \bar{a}$.

Dagsvik defines MEV (and calls it GEV) from functions A that satisfy limits and strong signs but $a-$homogeneity and not just 1-homogeneity. The degree of homogeneity is not important for us, as a MEV ARUM based on an $a-$homogeneous A is observationally equivalent to a MEV ARUM based on $A^{1/a}$, which is 1-homogeneous. Then Dagsvik proves the following theorem.

Theorem 25 *(Dagsvik, 1995) For any compact K and for any $\delta > 0$ there exists a MEV model with corresponding CDF $\tilde{F}\left(u - V\right)$ such that*

$$\sup_{V \in K} \left| P^{\tilde{F}}\left(j|V\right) - P^F\left(j|V\right) \right| < \delta.$$

The proof works by constructing a MEV model that tends to the true model at any $V \in K$. Since K is compact, convergence is uniform over K.

The MEV CDF is constructed as follows.

$$\tilde{F}\left(z\right) = \exp\left(-E\left(\sum_k \exp\left(\frac{V_k + \varepsilon_k - z_k}{a}\right)\right)^{a^2}\right)$$

and the corresponding choice probabilities are

$$P^{\tilde{F}}\left(j|V\right) = \frac{E\left(\exp\left(\frac{V_j+\varepsilon_j}{a}\right)\left(\sum_k \exp\left(\frac{V_k+\varepsilon_k}{a}\right)\right)^{a^2-1}\right)}{E\left(\sum_k \exp\left(\frac{V_k+\varepsilon_k}{a}\right)\right)^{a^2}}. \tag{50}$$

This expression tends to $P^F(j|V)$ as a tends to 0.

Combining this with Lemma 24 is sufficient to prove the following theorem.

Theorem 26 *For any compact K and for any $\delta > 0$ there exists a MEV model with corresponding CDF $\hat{F}(u - V)$ such that*

$$\sup_{V \in K} \left| P^{\hat{F}}(j|V) - P^F(j|V) \right| < \delta.$$

The approximating model is a cross-nested logit.

Proof. For each j, the numerator and the denominator of (50) can be approximated arbitrarily well by finite sums using Lemma 24. The approximations all use the same weights and mass points x_n^N. The approximation then has the form

$$P^{\hat{F}}(j|V) = \frac{\sum_{n=1}^{N} w_n^N \exp\left(\frac{V_j + x_{nj}^N}{a}\right)\left(\sum_k \exp\left(\frac{V_k + x_{nk}^N}{a}\right)\right)^{a^2-1}}{\sum_{n=1}^{N} w_n^N \left(\sum_k \exp\left(\frac{V_k + x_{nk}^N}{a}\right)\right)^{a^2}}.$$

The function $E\left(\sum_k \exp\left(\frac{V_k + \varepsilon_k - z_k}{a}\right)\right)^{a^2}$ is positive and bounded away from zero on K. Hence $P^{\hat{F}}(j|V)$ may be made arbitrarily close to $P^{\tilde{F}}(j|V)$ by choosing N sufficiently large. ■

Remark The model has the following associated MEV CPGF:

$$A\left(e^V\right) = \left[\sum_{n=1}^{N} w_n^N \left(\sum_k \exp\left(\frac{V_k + x_{nj}^N}{a}\right)\right)^{a^2}\right]^{\frac{1}{a}}, 0 < a < 1.$$

4 Multiple risk survival models

There is a close connection between random utility models and multiple risk survival models. This section explores this link.

Definition 27 *A survival model (SM) is a vector* $\exp(-\varepsilon) = \exp(-\varepsilon_1, ..., -\varepsilon_J)$ *of positive latent durations in* $\mathbb{R}^J$ *where* ε *has an absolutely continuous CDF* R *and* $E|\varepsilon| < \infty$. *The minimum duration* $\exp(-\varepsilon_0) = \min_j \exp(-\varepsilon_j)$ *is observed and induces an observable survival function* $S(t) = P(\exp(-\varepsilon_0) > t) = R(-\ln t, ..., -\ln t)$.

Remark When $\varepsilon_0 = \varepsilon_j$, say that exit occurs for the j'th cause. The cause is considered observable in some applications.

As in the RUM, the CDF R of latent durations may depend on observed and unobserved covariates. This notation is suppressed in the following. The CDF R is not identified in general without further assumptions (Kalbfleisch and Prentice, 2002). This motivates the following definition.

Definition 28 *If* $\exp(-\varepsilon) = \exp(-\varepsilon_1, ..., -\varepsilon_J)$ *is a SM and* $V \in \mathbb{R}^J$ *is a location vector, then* $\mathbf{T} = \exp(-(\mathbf{V} + \varepsilon))$, *where the minimum duration* $T_0 = \min_j T_j$ *is observed, is again a SM with survival function* $S(t|\mathbf{V}) = R(-(\mathbf{V} + \ln t))$. *A family of SM indexed by the location vector* $\mathbf{V}$ *is termed an additive survival model (ASM). Note that* R *is now identified from* $S(t|\mathbf{V})$ *since* $R(-\mathbf{V}) = S(1|\mathbf{V})$.

Theorem 29 *If* $\mathbf{T} = \exp(-(\mathbf{V} + \varepsilon))$ *is an ASM, then an associated CPGF defined by*

$$\ln G\left(e^{\mathbf{V}}\right) = -E(\ln T_0|\mathbf{V}) \tag{51}$$

exists and satisfies properties [G1]-[G4]. The probability of exit for cause j *is given by*

$$P(j|\mathbf{V}) = \frac{\partial \ln G\left(e^{\mathbf{V}}\right)}{\partial V_j}. \tag{52}$$

Conversely, if G *is a function satisfying properties [G1]-[G4], then there exists a ASM such that (51) and (52) hold.*

Proof. Let $\mathbf{T} = \exp\left(-\left(\mathbf{V}+\varepsilon\right)\right)$ be an ASM. Then

$$\begin{aligned} -E\left(\ln T_0|\mathbf{V}\right) &= -E\min_j\left(-\left(V_j+\varepsilon_j\right)\right) \\ &= E\max\left(V_j+\varepsilon_j\right). \end{aligned}$$

By Theorem 7, this exists and is a CPGF satisfying [G1]-[G4] with exit probabilities satisfying (52).

Conversely, let G satisfy [G1]-[G4]. Then by Theorem 7 there exists an ARUM $\mathbf{V}+\varepsilon$ such that G is an associated CPGF. From this define an ASM by $\mathbf{T} = \exp\left(-\left(\mathbf{V}+\varepsilon\right)\right)$. This ASM satisfies (51) and (52). ■

Definition 30 *An ASM has an exponential survival function if the hazard does not depend on t. The ASM is termed exponential in this case. The ASM has proportional hazard if the hazard factors as $\lambda(t|V) = A\left(e^V\right)\lambda_0(t)$.*

Theorem 31 *Consider an ASM $\mathbf{T} = \exp(-(\mathbf{V}+\varepsilon))$ with CPGF $G\left(e^{\mathbf{V}}\right) = \exp\left(-E\left(\ln T_0|\mathbf{V}\right)\right)$. Then ε is CAMS MEV if and only if the ASM is exponential. In this case, the hazard $A\left(e^{\mathbf{V}}\right)$ is equal to the CPGF $A\left(e^{\mathbf{V}}\right) = G\left(e^{\mathbf{V}}\right)$.*

Proof.

In case ε is CAMS MEV then the survival function is $S\left(t|\mathbf{V}\right) = \exp\left(-G\left(\exp\left(\mathbf{V}+\log t\right)\right)\right) = \exp\left(-tG\left(e^{\mathbf{V}}\right)\right)$ which is an exponential survival function. The hazard rate is given by $-\frac{\partial \ln S(t|\mathbf{V})}{\partial t} = G\left(e^{\mathbf{V}}\right)$, which does not depend on t.

Conversely, assume that the ASM is exponential with constant hazard given by $A\left(e^{\mathbf{V}}\right)$. Note that $S\left(t|\mathbf{V}\right) = S\left(1|\mathbf{V}+\ln t\right) = \exp\left(-A\left(e^{\mathbf{V}+\ln t}\right)\right)$, such that $S\left(t|\mathbf{V}\right) = \exp\left(-A\left(te^{\mathbf{V}}\right)\right) = \exp\left(-tA\left(e^{\mathbf{V}}\right)\right)$. Now,

$$R\left(\mathbf{V}\right) = S\left(1|-\mathbf{V}\right) = \exp\left(-A\left(e^{-\mathbf{V}}\right)\right),$$

with univariate marginals $R_{(j)}\left(V_j\right) = \exp\left(-e^{-V_j}A\left(0,...,0,1,0,...,0\right)\right)$ and inverses

$$R_{(j)}^{-1}\left(y_j\right) = \ln A\left(0,...,0,1,0,...,0\right) - \ln\left(-\ln y_j\right)$$

such that R has copula

$$\begin{aligned} C(y) &= \exp\left(-A\left(e^{-R_{(1)}^{-1}(y_1)}, ..., e^{-R_{(J)}^{-1}(y_J)}\right)\right) \\ &= \exp\left(-A\left(\frac{-\ln y_1}{A(1,0,...,0)}, ..., \frac{-\ln y_J}{A(0,...,0,1)}\right)\right). \end{aligned}$$

Finally, note that C is max-stable such that Lemma 18 implies that R is CAMS MEV and has the form $R(\mathbf{V}) = \exp\left(-G\left(e^{-\mathbf{V}}\right)\right)$. Hence $A\left(e^{-\mathbf{V}}\right) = G\left(e^{-\mathbf{V}}\right)$.

■

Theorem 32 *Any proportional hazard ASM is exponential.*

Proof. Let $\mathbf{T} = \exp(-(\mathbf{V}+\varepsilon))$ be the ASM and let R be the CDF of ε. Write the survivor function $P(T_0 \geq e^t) = R(-(t+\mathbf{V})) = \exp\left(-A\left(e^{\mathbf{V}}\right)\Lambda_0\left(e^t\right)\right)$. Then the integrated hazard $A\left(e^{\mathbf{V}}\right)\Lambda_0\left(e^t\right)$ satisfies

$$A\left(e^{\mathbf{V}}\right)\Lambda_0\left(e^s\right)\Lambda_0\left(e^t\right) = A\left(e^{\mathbf{V}}\right)\Lambda_0\left(e^{s+t}\right).$$

This implies that $\Lambda_0\left(e^s\right)\Lambda_0\left(e^t\right) = \Lambda_0\left(e^{s+t}\right)$, which in turn implies that $\Lambda_0\left(e^t\right) = e^t$, such that the hazard is constant. ■

5 Conclusion

This paper has contributed by completely characterizing the relationship between RUM and CPGF without imposing structure on utility. Furthermore, the paper has contributed by completely characterizing the subset of MEV ARUM in terms of CPGF and also the further subset of CAMS MEV ARUM in terms of CPGF. The subset of CAMS MEV ARUM consisting of cross-nested logit models is dense in the set of all RUM. The main results are extended to the case of multiple risk survival models. The present results generalize and subsume a range of previous contributions.

References

Antonelli, G. (1886) Sulla Teoria Matematica della Economia Politica *in* H. S. J. Chipman, L. Hurwicz (ed.), *Preferences, Utility, and Demand* Harcourt-Brace-Jovanovich New York.

Bates, J. (2003) Economic evaluation and transport modelling: theory and practice *in* K. W. Axhausen (ed.), *Moving through nets: the physical and social dimensions of travel* Elsevier.

Bierlaire, M. (2006) A theoretical analysis of the cross-nested logit model *Annals of operations research* **144**(1), 287–300.

Bierlaire, M., Bolduc, D. and McFadden, D. (2008) The estimation of Generalized Extreme Value models from choice-based samples *Transportation Research Part B: Methodological* **42**(4), 381–394.

Coles, S. (2001) *An Introduction to Statistical Modeling of Extreme Values* Springer-Verlag.

Dagsvik, J. (1995) How large is the class of generalized extreme value random utility models? *Journal of Mathematical Psychology* **39**(1), 90–98.

Daly, A. and Bierlaire, M. (2006) A General and Operational Representation of Generalised Extreme Value Models *Transportation Research Part B: Methodological* **40**(4), 285–305.

Daly, A. J. and Zachary, S. (1978) Improved multiple choice *in* D. A. Hensher and M. Q. Dalvi (eds), *Determinants of travel demand* Saxon House Sussex.

de Palma, A. and Kilani, K. (2007) Invariance of conditional maximum utility *Journal of Economic Theory* **132**(1), 137–146.

Diewert, E. (1974) Applications of Duality Theory *in* M. Intriligator and D. Kendrick (eds), *Frontiers of Quantitative Economics* North Holland Amsterdam.

Fisher, R. A. and Tippett, L. H. C. (1928) Limiting forms of the frequency distribution of the largest or smallest member of a sample *Mathematical Proceedings of the Cambridge Philosophical Society* **24**, 180–190.

Gnedenko, B. (1943) Sur La Distribution Limite Du Terme Maximum D'Une Série Aléatoire *The Annals of Mathematics* **44**(3), 423–453. Translated and reprinted in: Breakthroughs in Statistics, Vol. I, 1992, eds. S. Kotz and N.L. Johnson, Springer-Verlag, pp. 195-225.

Gorman, W. M. (1953) Community Preference Fields *Econometrica* **21**(1), 63–80.

Gumbel, E. (1958) *Statistics of Extremes* Columbia University Press New York.

Houthakker, H. (1950) Revealed Preference and the Utility Function *Economica* **17**, 159–174.

Ibáñez, J. N. and Batley, R. P. (2008) On the integrability conditions for discrete travel choice *in* A. for European Transport (ed.), *Proceedings of the European Transport Conference*.

Ibáñez, N. (2007) On the compatibility of probabilistic choice systems with random utility maximization *Working paper 592* ITS, University of Leeds.

Joe, H. (1997) *Multivariate Models and Dependence Concepts* Vol. 73 of *Monographs on Statistics and Applied Probability* Chapman and Hall/CRC.

Johnson, N., Kotz, S. and Balakrishnan, N. (1995) *Continuous univariate distributions* Vol. 2 2nd edn John Wiley New York.

Kalbfleisch, J. D. and Prentice, R. L. (2002) *The statistical analysis of failure time data* Vol. 360 of *Wiley series in probability and statistics* 2nd edn Wiley.

Koning, R. and Ridder, G. (2003) Discrete Choice and Stochastic Utility Maximization *Econometrics Journal* **6**, 1–27.

Lindberg, P. O., Eriksson, E. A. and Mattsson, L.-G. (1995) Invariance of achieved utility in random utility models *Environment & Planning A* **27**(1), 121–142.

Mas-Colell, A., Whinston, M. and Green, J. (1995) *Microeconomic Theory* Oxford Press Oxford.

McFadden, D. and Train, K. (2000) Mixed MNL Models for Discrete Response *Journal of Applied Econometrics* **15**(5), 447–470.

McFadden, D. (1974) The measurement of urban travel demand *Journal of Public Economics* **3**, 303–328.

McFadden, D. (1978) Modelling the choice of residential location *in* A. Karlquist *et al.* (ed.), *Spatial interaction theory and residential location* 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 application* MIT Press Cambridge, Mass.

McFadden, D. L. (2005) Revealed stochastic preference: a synthesis *Economic Theory* **26**(2), 245–264.

McFadden, D. and Richter, M. (1990) Stochastic Rationality and Revealed Stochastic Preference *in* M. R. J. Chipman, D. McFadden (ed.), *Preferences, Uncertainty, and Optimality: Essays in honor of Leonid Hurwicz* Westview Press Boulder and Oxford pp. 161–186.

Miller, K. S. and Samko, S. G. (2001) Completely monotonic functions *Integr. Transf. and Spec. Funct.* **12**(4), 389–402.

Nelsen, R. B. (2006) *An introduction to copulas* Springer Series in Statistics 2 edn Springer.

Richter, M. (1966) Revealed Preference Theory *Econometrica* **34**, 635–645.

Robertson, C. A. and Strauss, D. J. (1981) A characterization theorem for random utility variables *Journal of Mathematical Psychology* **23**(2), 184–189.

Samuelson, P. (1947) *Foundations of Economic Analysis* Harvard Press Cambridge.

Smith, T. E. (1984) A choice probability characterization of generalized extreme value models *Applied Mathematics and Computation* **14**(1), 35 – 62.

Sreehari, M. (2009) General max-stable laws *Extremes* **12**(2), 187–200.

Strauss, D. (1979) Some results on random utility models *Journal of Mathematical Psychology* **20**(1), 35–52.

Varian, H. (1993) *Microeconomic Analysis* Norton New York.

Vovsha, P. (1997) Cross-Nested Logit Model: an application to mode choice in the Tel-Aviv metropolitan area *Transportation Research Record* **1607**, 6–15.

Williams, H. (1977) On the formation of travel demand models and economic measures of user benefit *Environment and Planning* **9A**, 285–344.