A Note on Power Invariance in Random Utility Models.

P O Lindberg

April 17 2012

Abstract

Invariance in Random Utility (RU) Models is the property that the distribution of achieved utility is invariant across the alternatives chosen. In this note we study a generalization termed Power Invariance (PI). It generalizes the property, showed by Mattsson and Weibull, that invariance holds for an independent RU Model, where the cdf's of the random terms are positive powers of a given base cdf Φ. The power invariant copulas turn out to be the copulas of Multivariate Extreme Value distributions. We further show that, under natural regularity conditions, the copulas of the marginal distributions of a PI copula are power invariant.

Introduction

By invariance in random utility model, we mean that the distribution of achieved utility is invariant across the alternatives chosen. See below.

The study of *Invariance* in *Random Utility* (RU) *Models* originated in the works of Strauss (1979) and Robertson and Strauss (1981). They gave a characterization of the joint distribution of the random terms in *Additive* RU (ARU) Models, necessary and sufficient for the distribution of achieved utility to be independent of which alternative attains the maximum. Later Lindberg, Eriksson and Mattsson (1995), noted and corrected some errors in the proofs of Robertson and Strauss (1981), and gave an equivalent but different characterization. In this note, we take invariance one step further, by introducing the stronger property of power invariance.

In an RU Model, choice makers choose between alternatives in a finite *Universal Choice Set* $G =_{df} \{1,2,...,N\}$. With each alternative $i \in G$ is associated a random utility X_i, looked upon as the utility of alternative *i* for a randomly chosen choice maker. Hence, the probability of a randomly chosen choice maker to choose alternative *i* from a nonempty subset $I \subseteq G$ is postulated to be

$$p_I(i) =_{df} \Pr\{X_i > X_j, j \in I, j \neq i\}. \qquad (1)$$

It usually assumed that the probability of ties is zero, so the probabilities (1) sum up to 1

The utility distributions of the alternatives further typically depend on parameters, such as the costs of the alternatives, or the incomes of the choice makers when these are grouped into e.g. income classes.

A typical case in question is ARU models, where the utility X_i, has been endowed an "additive" structure:

$$X_i = v_i + U_i. \qquad (2)$$

Here the utility X_i, is decomposed into the sum of a deterministic *population value*, v_i, assumed known to the analyst (such as the cost of alternative *i*) and an *individual value*, U_i, assumed unknown to the analyst, and hence considered as random when studying choices by the choice makers. ARU models have become work horses in many areas of discrete probabilistic choice, such as choice of mode of transport (e.g. Ben-Akiva and Lerman, 1985), choice of residential location (e.g. McFadden, 1978), consumer brand choice, consumer theory and product differentiation (e.g. Anderson, de Palma and Thisse (1992)), and labor economics (e.g. Keane Todd and Wolpin 2010).

Let $\hat{X}_I$ denote the maximum achieved utility when choosing from the set *I*, i.e. $\hat{X}_I =_{df} \max_{i \in I}\{X_i\}$. Further let $\hat{X}_{i|I}$ denote the maximal utility conditioned upon alternative *i* being chosen out of *I*. When the choices are from the universal choice set *G*, we write just $\hat{X}$ and $\hat{X}_i$ for $\hat{X}_G$ and $\hat{X}_{i|G}$.

Let *F* be the joint *cdf* (cumulative distribution function) of $\mathbf{X} = (X_1, X_2,..., X_N)$. (We use bold face to denote vectors.) We say that *F* (and also the RU model in question) has the *invariance* *property*, if the distribution of $\hat{X}_i$ is independent of *i*, and that it has the *subset invariance property* if the distribution of $\hat{X}_{i|I}$ is independent of $i \in I$ for all $I \subseteq G$.

In an ARU model (2), the cdf of **X** depends on the vector **v** of population values, so let us denote it by $F_{\mathbf{v}}$, i.e. $F_{\mathbf{v}}(\mathbf{x}) =_{df} \Pr\{U_i + v_i \le x_i, i \in G\} = F_U(\mathbf{x}-\mathbf{v}) = F_{\mathbf{0}}(\mathbf{x}-\mathbf{v})$, where F_U is the cdf of $\mathbf{U} = (U_1, U_2, ..., U_N)$.

The invariance result proved (modulo the corrections by Lindberg et al. (1995)) by Robertson and Strauss (1981) says that for an ARU Model, all the $F_{\mathbf{v}}$ have the invariance property, if and only if the cdf F_U has the form

$$F_U(\mathbf{x}) = \phi(H(e^{-\mathbf{x}}) =_{df} \phi(H(e^{-x_1}, e^{-x_2}, ..., e^{-x_N}), \qquad (3)$$

where *H* is a linearly homogenous function and ϕ a scalar function. (In (3), and in the sequel, we use the convention, that a function of one variable applied to a vector is the vector of function values of the vector components, i.e. $f(\mathbf{x}) =_{df} (f(x_1), f(x_2), ..., f(x_n))$.) This class contains among others the GEV (*Generalized Extreme Value*) distributions (McFadden, 1978), for which $\phi(x) = e^{-x}$.

Recently, Li (2011) as well as Mattsson & Weibull (2010) have noted that, in an RU Model with <u>independent</u> random utitilities X_i (an *independent* RU Model), each with cdf of the form Φ^{α_i} with $\alpha_i > 0$, for a given "*base*" cdf Φ, the choice probabilities have the well-known "Luce form":

$$p_I(i) =_{df} \Pr\{X_i > X_j, j \in I, j \ne i\} = \frac{\alpha_i}{\sum_{j \in I} \alpha_j}, \qquad (4)$$

(cf. Luce (1959)).

Lindberg (2012a) has further shown a converse to this result, namely that if a class of RU Models allows for "non-uniform expansion" of the choice set, and if it has choice probabilities of the Luce form (4) then the random variables X_i must have cdf's of the form Φ^{α_i} for some base cdf Φ, possibly modulo some freakish cases. This result proves a conjecture by Luce and Suppes (Luce & Suppes, 1965, p 338-339).

Mattsson and Weibull (2010, Prop.2) further show that an independent RU Model with cdf's of the form Φ^{α_i} with positive α_i for a given base cdf Φ, has the *invariance property*. It is also easy to see that their proof carries over to *subset* invariance.

Thus, these distributions, termed *Mattsson-Weibull-Li* (MWL) distributions by Lindberg (2012a), have many interesting properties. Let us state the last result separately for future reference:

Lemma 1. If *F* is the cdf of a MWL-distribution, then *F* has the subset invariance property. □

Further results are given in Theorems 1 and 1' below. In the present paper we study a possible generalization of the invariance property of MWL-distributions to dependent variables.

When we, in the sequel talk about more general distributions with the invariance property, we will sharpen the definition to say that a cdf $F: C \subseteq \Re^N \to (0,1)$ has the *invariance property* if a RU Model with random utilities distributed according to *F*, has the above mentioned invariance property, and moreover, each alternative has a positive probability of being chosen. This sharpened definition is needed to rule out freakish behavior of alternatives that have no chance of being chosen.

1. *Preamble*

The MWL distributions have further interesting properties, as showed by the following converse (part of their Thm. 2) by Mattsson, Weibull and Lindberg (2011) to the above mentioned Prop. 2 of Mattsson and Weibull (2010).

Theorem. 1. Let $X_1, X_2,.., X_N$ be statistically independent with cdf's $F^{(i)} : E \to (0,1)$, for some open set $E \subseteq \Re$. If all conditional random variables $\hat{X}_i$ have the same probability distribution (i.e. the joint cdf *F* has the invariance property), then for any *i, j* there is an *a*>0, such that $F^{(i)} = (F^{(j)})^a$. □

Thus, in the independent case, invariance implies MWL distributions, if the cdf's have the same domain. For the differentiable case, we will give a sharpened version of this result below.

Let us first note that, in general under the invariance property, all random variables must have the same sup (supremum) of their supports

Lemma 2. Suppose that $F : \Re^n \to (0,1)$ is a differentiable cdf with the invariance property, and that its marginal distributions have cdf's with supports $[c_i, d_i]$. Then all d_i are equal.

Proof: Suppose the contrary. Then for some subset $\overline{G}$ of *G*={1,…, *N*}, $d_i \equiv \overline{d} =_{df} \max_{k \in G} d_k$, and for the rest of the indices, $\tilde{G} =_{df} \{i \in G | i \notin \overline{G}\}$, we have $d_i \le \tilde{d} =_{df} \sup_{k \in \tilde{G}} d_k < \overline{d}$. It is obvious that for $i \in \overline{G}$, $\hat{X}_i$ must have mass in $(\tilde{d}, \overline{d}]$. It is equally obvious that for $i \in \hat{G}$, $\hat{X}_i$ cannot have mass in $(\tilde{d}, \overline{d}]$. This contradicts that all $\hat{X}_i$ have the same distribution. □

Comment: I would have liked to prove that the c_i are equal, too, but so far I have not managed to do that. It is however valid in the differentiable independent case. See Thm. 1' below.

We are now in a position to sharpen the result in Thm. 1 in one direction. Then we will need the following result from Lindberg (2012b). To this end we define the *diagonal* $D = \{\mathbf{x} | \mathbf{x} = \lambda \mathbf{1}, \lambda \in \Re\}$ in $\Re^N$, where $\mathbf{1}$ denotes the *one-vector* $\mathbf{1} = (1,1,\ldots,1)$.

Theorem 2. Let $F : E \subseteq \Re^N \to (0,1)$ be a cdf that is differentiable on $D \cap E$, with locally bounded gradient. Consider a RU model with random term distributed according to *F*. Then the cdf $\hat{F}^{(i)}$ of $\hat{X}_i$ fulfills

$$\hat{F}^{(i)}(x) =_{df} \Pr\{\hat{X}_i \le x\} = \Pr\{X_j \le X_i, j \ne i \text{ and } X_i \le x\} = \frac{1}{p_i} \int_{-\infty}^{x} F_i(s\mathbf{1}) ds, \tag{5}$$

where $p_i =_{df} p_I(i)$ and F_i denotes the partial derivative of *F* w.r.t x_i. In particular

$$p_i = \int_{-\infty}^{\infty} F_i(s\mathbf{1}) ds. \tag{6}$$ □

As a corollary we have the following useful result.

Corollary 1. Let $F: E \subseteq \Re^N \to (0,1)$ be a cdf, differentiable on $D \cap E$, with locally bounded gradient, and consider a RU model with random term distributed according to *F*. Then *F* has the invariance property if and only if, for all *i* there are $a_i > 0$ such that

$$F_i(s\mathbf{1}) = a_i F_1(s\mathbf{1}) \text{ a.e. on } \{s \in \Re | s\mathbf{1} \in E\}. \qquad (7)$$

Remark: We might as well extend *F* to $\Re^N$, but then it need not be differentiable on the boundary of $D \cap E$.

Proof: *F* has the invariance the invariance property if and only if $\hat{F}^{(i)}(x) \equiv \hat{F}^{(1)}(x)$ for all *i*. I.e., by (5) if and only if, for all *i*, and *x*,

$$\frac{1}{p_i} \int_{-\infty}^{x} F_i(s\mathbf{1}) ds = \frac{1}{p_1} \int_{-\infty}^{x} F_1(s\mathbf{1}) ds. \qquad (8)$$

If (8) is valid, we get by differentiation, that for all *i*

$$\frac{1}{p_i} F_i(x\mathbf{1}) = \frac{1}{p_1} F_1(x\mathbf{1}) \text{ a.e.,} \qquad (9)$$

i.e. (7) with $a_i = p_i / p_1$

Conversely, if (7) is valid for all *i*, then we get by integration from $-\infty$ to *x*: $\int_{-\infty}^{x} F_i(s\mathbf{1}) ds = a_i \int_{-\infty}^{x} F_1(s\mathbf{1}) ds$. Letting $x = \infty$, we get $a_i = p_i / p_1$, by (6), whence (8) follows.

□

From the corollary it is clear that whether a given cdf *F* has the invariance property, only depends on its behavior around the diagonal. Thus, invariance does not necessarily imply subset invariance. We will return to subset invariance in the last section.

Now to the sharpened version of Thm. 1.

Theorem. 1'. (Sharpened version of Thm.1. for the differentiable case) Let $X_1, X_2, .., X_N$ be statistically independent with differentiable cdf's $F^{(i)}$, with supports $\overline{E}_i = [c_i, d_i] \subseteq \Re$. If all conditional random variables $\hat{X}_i$ have the same probability distribution, then for all *i*, there is an *a*>0, such that $F^{(i)} = (F^{(1)})^a$. This implies in particular that all $[c_i, d_i]$ are equal.

Proof: Assume we have invariance. Then by Lemma 2, all $d_i = \overline{d} = \max_k d_k$.

Let $\overline{c} =_{df} \max_i c_i$, and assume (possibly by renumbering) that $c_1 = \overline{c}$. Then all $F^{(i)}$ are positive on $(\overline{c}, \overline{d})$.

Let *j* be arbitrary >1. By Corollary 1, we must have

$$\frac{1}{p_1} F_1(s\mathbf{1}) = \frac{1}{p_j} F_j(s\mathbf{1}) \text{ a.e. on } (\overline{c}, \overline{d}) \qquad (10)$$

Let $f^{(i)}$ denote the derivative of $F^{(i)}$.

Now $F(\mathbf{x}) = \Pr\{\mathbf{X} \le \mathbf{x}\} = \{\text{by independen ce}\} = \prod_i F^{(i)}(x_i)$.

Thus $F_i(\mathbf{x}) = f^{(i)}(x_i)\prod_{j\neq i} F^{(j)}(x_j)$, and by (10) we get

$$\frac{1}{p_1} f^{(1)}(x)\prod_{i\neq 1} F^{(i)}(x) = \frac{1}{p_j} f^{(j)}(x)\prod_{i\neq j} F^{(i)}(x) \text{ a.e., or} \qquad (11)$$

Since all $F^{(i)}$ are positive on $(\bar{c},\bar{d})$, we can divide by $F(s\mathbf{1}) = \prod_i F^{(i)}(s)$ in (11), to get

$$\frac{1}{p_1} f^{(1)}(x)/F^{(1)}(x) = \frac{1}{p_j} f^{(j)}(x)/F^{(j)}(x) \text{ a.e. on } (\bar{c},\bar{d}). \qquad (12)$$

Integrating (12) from $x \in (\bar{c},\bar{d})$ to $\bar{d}$, we get

$$\frac{1}{p_1}[0 - \ln(F^{(1)}(x)) = \frac{1}{p_j}[0 - F^{(j)}(x)] \text{ on } (\bar{c},\bar{d}), \text{ or}$$

$$p_j \ln(F^{(1)}(x)) = p_1 \ln(F^{(j)}(x)) \text{ on } (\bar{c},\bar{d}], \text{ or}$$

$$(F^{(1)}(x))^{p_j} = (F^{(j)}(x))^{p_1}, \text{ i.e.}$$

$$F^{(j)}(x) = (F^{(1)}(x))^{p_j/p_1}. \qquad \square$$

2. *Main result*

In our endeavor it is fruitful to look at the concept of *copulas*. A copula is a cdf *F* on $[0,1]^N$, with uniform [0,1] marginal distributions (Nelsen, 2006, Def. 2.10.6). Copulas are tools for introducing dependence among random variables in a controlled way. If *F* is a copula, and $\{\Phi^{(i)}\}_{i=1}^N$ is a set of univariate cdf's, then $F(\Phi^{(1)},\Phi^{(2)},...,\Phi^{(N)})$ is a cdf with marginals $\Phi^{(i)}$. Conversely, by Sklar's Theorem (Nelsen (2006), Thm. 2.10.9), if *F* is a cdf with univariate marginals $\Phi^{(i)}$, then there is a copula C_F, such that $F(x_1, x_2,..., x_N) = C_F(\Phi^{(1)}(x_1), \Phi^{(2)}x_2),...,\Phi^{(n)}(x_N))$.

In the independent differentiable case, when we have invariance, all marginal distributions have the same support. In generalizing Thms. 1 and 1' to the dependent case, we will assume the same.

We will further assume that the dependence is introduced by a given differentiable copula $C:[0,1]^N \to [0,1]$, which we assume has the invariance property. Then transforming the independent variables with some strictly monotone and differentiable cdf $\Phi: E \subseteq \Re \to [0,1]$, we retain the invariance property, i.e. the cdf *F* defined by $F(\mathbf{x}) =_{df} C(\Phi(x_1), \Phi(x_2),..., \Phi(x_n))$ has the invariance property. To see this note e.g. that for any real *x*,
$F_i(x, x,..., x) = C_i(\Phi(x), \Phi(x),..., \Phi(x))\Phi'(x) = \{\text{by Cor.1, since } C \text{ has the invariance property}\}$
$a_i C_1(\Phi(x), \Phi(x),..., \Phi(x))\Phi'(x) = a_i F_1(x, x,..., x)$, for some $a_i > 0$.

Thus *F* has the invariance property by Cor. 1.

In the MWL-case, for a given base cdf Φ and powers a_i, the marginal distributions are Φ^{a_i}. Thus the corresponding copula is $\bar{C}(\mathbf{x}) =_{df} x_1 x_2 ... x_n$, the "*independence copula*". (Plugging in any cdf $\Phi^{(i)} : E_i \to (0,1)$ into the arguments of $\overline{C}$, we get a cdf with independent random variables.) $\overline{C}$ obviously has the invariance property (by Lemma 1). By Lemma 1, it has the further property that for any $\mathbf{a} \in \Re^n_{++}$ the cdf $\overline{F}(\mathbf{x}) =_{df} \overline{C}(\mathbf{x}^{\mathbf{a}}) = x_1^{a_1} x_2^{a_2} \ldots x_n^{a_n}$ also has the invariance property. Then, for a given cdf $\Phi : E \subseteq \Re \to [0,1]$, the cdf $\overline{F}(\Phi(\mathbf{x})) = \overline{C}(\Phi(\mathbf{x})^{\mathbf{a}}) = \Phi^{a_1}(x_1)\Phi^{a_2}(x_2)\ldots\Phi^{a_n}(x_n)$ has the invariance property, as noted above, which is just what we want.

Therefore, let us say that a copula C has the *power invariance* (PI) *property* if for any $\mathbf{a} = (a_1, a_2, \ldots, a_n) \in \Re^n_{++}$ the cdf $\overline{F}_a(\mathbf{x}) =_{df} C(\mathbf{x}^{\mathbf{a}})$ has the invariance property. In particular the independence copula $\overline{C}$ above has the power invariance property.

In proving the main result, we need the following result from Lindberg, Eriksson and Mattsson (2002).

Theorem 3. (Thm. 1 in Lindberg et al. (2002).) Let the continuous function *f* be defined on a weakly solid cone *K* in $\Re^n$, and continuously differentiable on $\operatorname{int} K_0$. Suppose that *f* is nowhere ray constant. Then *f* is homothetic if and only if it has ray parallel gradients. □

There are some terms in the statement that need explanation. A set *E* is termed *weakly solid* if its interior $\operatorname{int} E$ is non-empty and connected, and contains *E* in its closure. E_0 denotes the set $E_0 = \{\mathbf{x} \in E \mid \mathbf{x} \neq \mathbf{0}\}$. We also say that the function *f* is *nowhere ray constant* on a cone *K* in $\Re^n$, if for all nonzero $\mathbf{x} \in K$, there is some $\lambda > 0$ such that $f(\lambda\mathbf{x}) \neq f(\mathbf{x})$. Further, the function *f* on *K* is said to be *homothetic* (Shephard (1953)) if $f(\mathbf{x}) = h(A(\mathbf{x}))$, where *h* is strictly monotone, and *A* is linearly homogenous. Finally we say that the function *f* defined on some cone *K* in $\Re^n$ has *ray parallel gradients* if there is a function $k : \Re_{++} \otimes C_0 \to \Re$ such that

$$\nabla f(\lambda\mathbf{x}) = k(\lambda, \mathbf{x})\nabla f(\mathbf{x}), \tag{13}$$

for all $\lambda > 0$ and $\mathbf{x} \in K_0$.

Now to the main result.

Theorem 4. A continuously differentiable copula *C*, with non-vanishing gradient, has the PI property if and only if it can be written in the form

$$C(\mathbf{x}) = \exp(-A(-\ln \mathbf{x})), \tag{14}$$

for some differentiable linearly homogenous function $A : \Re^N_{++} \to (0,\infty)$, with $A(\mathbf{u}_i) = 1$, where $\mathbf{u}_i$ is the *i*-th coordinate unit vector.

Remark: Strictly speaking, in our setup $A(\mathbf{u}_i)$ is not defined since *A* is only defined on $\Re^N_{++}$. $A(\mathbf{u}_i)$ should instead be interpreted as the limit of $A(\mathbf{x})$ when $\mathbf{x} \in \Re^N_{++}$ tends to $\mathbf{u}_i$. We will return to this aspect in the last section.

Proof: <u>If part</u>: Suppose

$$F(\mathbf{x}) = C(\mathbf{x}^{\mathbf{a}}) = \exp(-A(-\ln \mathbf{x}^{\mathbf{a}})) = \exp(-A(-\mathbf{a} * \ln \mathbf{x})) \tag{15}$$

for some differentiable linearly homogenous $A : \Re^N_{++} \to [0,\infty)$, and for an arbitrary $\mathbf{a} \in \Re^N_{++}$. (Here, and in the sequel, we use * to denote component-wise multiplication between vectors, i.e. $\mathbf{x}*\mathbf{y} =_{df} (x_1y_1, x_2y_2,,...,x_Ny_N)$. Similarly we will use / also for component-wise division of vectors, i.e. $\mathbf{x}/\mathbf{y} =_{df} (x_1/y_1, x/_2\, y_2,,...,x_N/y_N)$. Finally, in line with these conventions, we also use the notation $\mathbf{x}^{\mathbf{y}} =_{df} (x_1^{y_1}, x_2^{y_2},..,x_N^{y_N})$) Then

$F_i(\mathbf{x}) = \exp(-A(-\mathbf{a}*\ln\mathbf{x}))(-A_i(-\mathbf{a}*\ln\mathbf{x}))(-a_i/x_i)$. Thus for an arbitrary $s \in (0,1)$,

$$F_i(s\mathbf{1}) = \exp(-A((-\ln(s\mathbf{1}))*\mathbf{a}))(A_i((-\ln(s\mathbf{1}))*\mathbf{a}))(a_i/s) =$$

$$\exp(-A((-\ln s)*\mathbf{a}))(A_i((-\ln s)*\mathbf{a}))(a_i/s) =$$

$$\{\text{using the homogeneity of } A \text{ and } A_i \text{ to the 1st and 0-th order}\}$$

$$= \exp((\ln s)A(\mathbf{a}))A_i(\mathbf{a})(a_i/s) = s^{A(\mathbf{a})}A_i(\mathbf{a})a_i/s = \frac{a_i}{a_1}s^{A(\mathbf{a})}A_i(\mathbf{a})a_1/s = \frac{a_i}{a_1}F_i(s\mathbf{1}), \text{ for } s \in (0,1),$$

proving invariance by Cor. 1.

Only if part: Suppose the differentiable copula C has the PI property, i.e. for each $\mathbf{a} \in R^N_{++}$, the cdf $F_a(\mathbf{x}) =_{df} C(\mathbf{x}^{\mathbf{a}})$ has the invariance property.

We can now transform the argument of F_a without loosing the invariance. Thus let $\mathbf{x} = e^{\mathbf{y}}$, i.e. $x_i = e^{y_i}$ for $y_i \in [-\infty,0]$, and let $\overline{F}(\mathbf{y}) =_{df} C(e^{\mathbf{y}})$. (Note that e^{y_i} is the cdf of an exponential distribution on the negative half-axis.) Then, the cdf $\overline{F}_{\mathbf{a}}(\mathbf{y}) =_{df} C(e^{\mathbf{a}*\mathbf{y}}) = \{\text{corresp. to } C(\mathbf{x}^{\mathbf{a}}) = C((e^{\mathbf{y}})^{\mathbf{a}}) = C(e^{\mathbf{a}*\mathbf{y}})\} = \overline{F}(\mathbf{a}*\mathbf{y})$ has the invariance property. We get for arbitrary *i*, and $s \in (-\infty,0)$

$$\overline{F}_{\mathbf{a},i}(s\mathbf{1}) =_{df} \frac{\partial \overline{F}_{\mathbf{a}}}{\partial x_i}(s\mathbf{1}) = \frac{\partial}{\partial x_i}\overline{F}(\mathbf{a}*s\mathbf{1}) = a_i\overline{F}_i(\mathbf{a}*s\mathbf{1}) = a_i\overline{F}_i(s\mathbf{a}). \qquad (16)$$

On the other hand, since $\overline{F}_{\mathbf{a}}$ has the invariance property, there are $b_i > 0$, such that

$$\overline{F}_{\mathbf{a},i}(s\mathbf{1}) = b_i\overline{F}_{\mathbf{a},1}(s\mathbf{1}) = \{\text{by (16)}\} = b_ia_1\overline{F}_1(s\mathbf{a}). \qquad (17)$$

k (16) and (17) imply that

$$\overline{F}_i(s\mathbf{a}) = b_ia_1\overline{F}_1(s\mathbf{a})/a_i, \text{ for } s \in (-\infty,0),$$

which says that along the ray $\{\mathbf{y} \in \Re^n_{--} \,|\, \mathbf{y} = s\mathbf{a}, s < 0\}$, the gradients of $\overline{F}$ are parallel, i.e. since $\mathbf{a}$ was arbitrary, that $\overline{F}$ has ray parallel gradients. Further note that since C is differentiable with non-vanishing gradient, the same is true for $\overline{F}$. Moreover we have $\nabla\overline{F} \geq \mathbf{0}$, since $\overline{F}$ is component-wise non-decreasing.

If now $\overline{F}$ is not nowhere ray constant, then there are $\overline{\mathbf{y}} \in -\Re^N_{++}$ and $\overline{\lambda} > 1$ such that $\overline{F}(\overline{\mathbf{y}}) = \overline{F}(\overline{\lambda}\overline{\mathbf{y}})$. But then $\overline{F}(\lambda\overline{\mathbf{y}})$ is constant for $1 \leq \lambda \leq \overline{\lambda}$, which implies that $\nabla\overline{F}(\overline{y}) \cdot \overline{y} = \mathbf{0}$, which in turn implies $\nabla\overline{F}(\overline{\mathbf{y}}) = \mathbf{0}$, since $\overline{\mathbf{y}} \in -\Re^n_{++}$ and $\nabla\overline{F} \geq \mathbf{0}$. Thus $\overline{F}$ is nowhere ray constant.

Therefore by Thm. 2, $\overline{F}$ is homothetic, i.e. there is a strictly monotone function *h*, and a linearly homogenous function *A*, such that $\overline{F}(\mathbf{y}) = h(A(\mathbf{y}))$.

A is so far undetermined in sign. A change in sign can be incorporated into h. Take an arbitrary $\bar{\mathbf{y}} \in -\Re_{++}^N$. As s goes from ∞ to 0, $s\bar{\mathbf{y}}$ goes from $\infty \cdot \bar{\mathbf{y}}$ to 0. At the same time $\bar{F}(s\bar{\mathbf{y}}) = h(A(s\bar{\mathbf{y}}))$ goes from 0 to 1. Thus we cannot have $A(\bar{\mathbf{y}}) = 0$. Assume, WLOG, that $A(\bar{\mathbf{y}}) > 0$. Then, by homogeneity, $A(s\bar{\mathbf{y}})$ goes from ∞ to 0, as s goes from ∞ to 0. Hence h is defined on $(0,\infty)$ and is strictly decreasing from 1 to 0.

Thus To get a more natural behavior we will redefine h to $\bar{h}(x) = h(-x)$. Then we get $\bar{F}(\mathbf{y}) = \bar{h}(-A(\mathbf{y}))$, where $\bar{h}$ is defined on $(-\infty,0)$, and is strictly increasing from 0 to 1.

A is defined on $(-\infty,0)^N$, which is a bit awkward. By introducing $\bar{A}(\mathbf{y}) =_{df} A(-\mathbf{y})$, we get a linearly homogenous function defined on $(0,\infty)^N$ instead. Thus we arrive at

$\bar{F}(\mathbf{y}) = \bar{h}(-\bar{A}(-\mathbf{y}))$, where $\bar{A}$ is defined on $(0,\infty)^N$.

Now take a fixed $\mathbf{y} \in \Re_{--}^N$ such that $\bar{A}(-\mathbf{y}) = 1$. Then, for $\lambda > 0$, $\bar{F}(\lambda\mathbf{y}) = \bar{h}(-\bar{A}(-\lambda\mathbf{y})) = \bar{h}(-\lambda\bar{A}(-\mathbf{y})) = \bar{h}(-\lambda)$. Differentiating, we get $\bar{h}'(-\lambda) = -\nabla\bar{F}(\lambda\mathbf{y}) \cdot \mathbf{y}$. But as noted above, $\nabla\bar{F}(\lambda\mathbf{y}) \cdot \mathbf{y}$ cannot be zero unless $\nabla\bar{F}(\lambda\mathbf{y}) = \mathbf{0}$, which is out-ruled. Thus $\bar{h}$ is differentiable with non-vanishing (i.e. positive) derivative, and hence strictly increasing.

Therefore, it has an inverse $\bar{h}^{-1}$, that is also strictly increasing and differentiable. Thus we get $\bar{h}^{-1}(\bar{F}(\mathbf{y})) = -\bar{A}(-\mathbf{y})$, which implies that $\bar{A}$ is differentiable.

Now let's transform back to the original copula C. We had $\bar{F}(\mathbf{y}) = C(e^{\mathbf{y}})$, whence $C(\mathbf{x}) = \bar{F}(\ln(\mathbf{x})) = \bar{h}(-\bar{A}(-\ln(\mathbf{x})))$.

For C to be a copula we need the univariate marginals to be uniform (0,1). To check the marginals, we let all x_j except one, x_i say, which is held at x, tend to their upper bounds 1. We get, for the i-th marginal,

$$C^{(i)}(x) = \lim_{x_j \to 1, j \neq i} C(x_1, x_2, ..., x_{i-1}, x, x_{i+1}, .. x_N) =$$

$$\lim_{x_j \to 1, j \neq i} \bar{h}(-\bar{A}(-\ln(x_1, ..., x_{i-1}, x, x_{i+1}, .. x_n))) = \bar{h}(-\bar{A}((0, ..., -\ln x, ..., 0)))$$

$$= \{\text{by homogeneity}\} = \bar{h}(\ln(x)\bar{A}(\mathbf{u}_i))$$

But $C^{(i)}(x) = x$. Thus we get $\bar{h}(\ln(x)\bar{A}(\mathbf{u}_i)) = x$. Let $y = \bar{A}(\mathbf{u}_i)\ln(x)$, i.e. $y / \bar{A}(\mathbf{u}_i) = \ln(x)$, or $x = \exp(y / \bar{A}(\mathbf{u}_i))$, then $\bar{h}(y) = \exp(y / \bar{A}(\mathbf{u}_i))$. This implies that all $\bar{A}(\mathbf{u}_i)$ are equal, and by redefining $\bar{h}$, we may assume that $\bar{A}(\mathbf{u}_i) = 1$ for all i. Then we get $\bar{h}(y) = \exp(y)$, and we have finally arrived at the desired representation

$C(\mathbf{x}) = \exp(-\bar{A}(-\ln(\mathbf{x})))$, with $\bar{A}(\mathbf{u}_i) = 1$, for all i. □

M Fosgerau suggested proving Thm. 4 as a corollary to Thm. 6.10 in Joe (1997). This is possible, but it turns out that Joe, in the proof follows Robertson and Strauss (1981). Thus he makes essentially the same error as Robertson and Strauss, an error, which was noted and corrected in Lindberg et al. (1995). (This makes the third error of this type that we have noted and corrected. In Lindberg et al. (2002), we detected and corrected a similar error in Lau (1969). Instead, it is possible to prove Joe's Thm. 6.10 using our Thm. 3.

As final result we state the following simple observations, some of which are given without proof:

Observation 1. The distributions (15) are the distributions obtainable from a *Power Invariance Copula,* $C(\mathbf{x}) = \exp(-A(-\ln(\mathbf{x})))$, by introducing powers a_i of the arguments. If we further plug in an arbitrary base cdf Φ in the arguments of (16), we arrive exactly at the distributions with cdf's of the form

$$F(\mathbf{x}) = \exp(-A(-\mathbf{a} * \ln(\Phi(\mathbf{x}))))$$

introduced by Mattsson et al (2011). □

Observation 2. The copula of any GEV distribution is a *PI Copula*, i.e. of the form

$$C(\mathbf{x}) = \exp(-A(-\ln(\mathbf{x})) \qquad (14)$$

and vice versa.

(To see this note that a GEV-distributions has cdf F_H given by $F_H(\mathbf{x}) = \exp(-H(e^{-\mathbf{x}}))$, where *H* is linearly homogenous, and differentiable. F_H has marginals $F_H^{(i)}(x) = \lim_{\substack{x_j \to \infty, j \neq i \\ x_i = x}} \exp(-H(e^{-\mathbf{x}})) = \exp(-e^{-x} H(\mathbf{u}_i)) = \{\text{with } a_i =_{df} H(\mathbf{u}_i)\} = \exp(-a_i e^{-x})$,

i.e. Gumbel distributions. Thus to transform F_H to a copula, we need to transform the arguments by the inverses of the functions $y_i = \exp(-a_i e^{-x_i})$, or $x_i = -\ln(-(\ln y_i)/a_i))$. We get the copula
$C_H(\mathbf{y}) = \exp(-H(\exp(\ln(-(\ln \mathbf{y})/\mathbf{a})))) = \exp(-H(-\ln(\mathbf{y})/\mathbf{a})) = \exp(-A(-\ln(\mathbf{y})))$, where $A(\mathbf{x}) = H(\mathbf{x}/\mathbf{a})$. *A* obviously is linearly homogenous, and $A(\mathbf{u}_i) = H(\mathbf{u}_i/\mathbf{a}) = H(\mathbf{u}_i/a_i) = H(\mathbf{u}_i)/a_i = 1$, i.e. C_H is of the form (14).

Conversely, assume that $C(\mathbf{y}) = \exp(-A(-\ln(\mathbf{y}))$ is an arbitrary PI copula. Inserting Gumbel cdf's $y_i = \exp(-a_i e^{-x_i})$ in the arguments of this copula (18), we get a GEV distribution, with $H(\mathbf{x}) =_{df} A(a * \mathbf{x})$, which is homogenous.) □

Observation 3. The PI copulas of the form (14) form the set of copulas for Multivariate Extreme Value MEV) distributions (Joe, (1997), § 6.2)

(Too se this note with Joe (p. 173) that an MEV copula *C* must satisfy $C(\mathbf{x}^t) = C^t(\mathbf{x})$. With $D(\mathbf{y}) =_{df} C(e^{-\mathbf{y}})$ (i.e. $\mathbf{y} = e^{-\mathbf{x}}$ and $\mathbf{x} = -\ln \mathbf{y}$), Joe notes (p. 174) that $A =_{df} \ln D$ is linearly homogenous. Thus $D(\mathbf{y}) = e^{-A(\mathbf{y})}$, and $C(\mathbf{x}) = D(-\ln \mathbf{x}) = \exp(-A(-\ln \mathbf{x}))$ for a linearly homogenous *A*.) □

3. *Subset invariance and power invariance*.

We now return to subset invariance, and in particular subset invariance in the case of power invariance. Since the invariance properties are inherited by the copulas, we may as well study copulas. So let *C* be an arbitrary power invariant copula, i.e. by (14)

$$C(\mathbf{x}) = \exp(-A(-\ln \mathbf{x})), \qquad (15)$$

Here, A is defined on $\mathfrak{R}_{++}^N$, implying that we have $x_i \in (0,1)$ in (15). Suppose we want to look at invariance for the subset I of G. By renumbering, we may assume $I = \{1,2,\ldots,n\}$. The marginal distribution of $\mathbf{X}_I = (X_1, X_2,\ldots, X_n)$ as a function of $\mathbf{x}_I = (x_1, x_2,\ldots x_n)$ is given by

$$F^{(I)}(\mathbf{x}_I) =_{df} \lim_{\substack{y_k \to 1.k>n \\ \mathbf{y}_I = \mathbf{x}_I}} C(\mathbf{y}) = \lim_{\substack{x_k \to 1.k>n \\ \mathbf{y}_I = \mathbf{x}_I}} \exp(-A(-\ln \mathbf{y})) = \exp(-A(-\ln(\mathbf{x}_I),0,\ldots 0)). \qquad (19)$$

Here again, $A(-\ln \mathbf{x}_I,0,\ldots 0))$ has to be interpreted as a limit. This limit exists, since the LHS limit of (19) is well-defined and equal to the RHS.

Now, for any $\mathbf{x} \in \mathfrak{R}_+^N$ let $I(\mathbf{x}) = \{i | x_i > 0\}$ be the index set of positive components of $\mathbf{x}$, and conversely $\tilde{I}(\mathbf{x}) = \{i | x_i = 0\}$ the index set of zero components. Then we can extend the definition of A to $\mathfrak{R}_+^N$ by defining

$$\overline{A}(\mathbf{x}) =_{df} \lim_{\substack{y_i \to 0, i \in \tilde{I}(\mathbf{x}) \\ y_i = x_i, i \in I(\mathbf{x})}} A(\mathbf{y}) \qquad (20)$$

The limits in (20) are well-defined for the same reasons as those in (19). Let $x_I =_{df} (x_i)_{i \in I}$, and let $\bar{x}_I \in \mathfrak{R}^N$ have components equal to those of x_I for $i \in I$, and equal to zero for $i \notin I$. Now, we can define the restriction $\overline{A}^{(I)}$ of $\overline{A}$ to any coordinate space I, i.e. the subspace where only $x_i, i \in I$ are non-zero:

$$\overline{A}^{(I)}(\mathbf{x}_I) =_{df} \overline{A}(\overline{\mathbf{x}}_I).$$

Using (19), we then have:

Lemma 3. The cdf of the marginal distribution in the I-coordinate space, of a PI copula (15) can be written

$$F^{(I)}(\mathbf{x}_I) = \exp(-\overline{A}^{(I)}(-\ln(\mathbf{x}_I))),$$

Where $\overline{A}^{(I)}$ is the restriction to the I-coordinate space of the extension $\overline{A}$ of A to $\mathfrak{R}_+^N$.

□

As a consequence we have, by Thm. 4:

Corollary 2. If the restriction $\overline{A}^{(I)}$ to the I-coordinate space, of the extension $\overline{A}$ of A, is differentiable with locally bounded gradient in that space, then the copula of the marginal distribution in I-space, of the PI copula (15) is power invariant.

As a consequence, we have:

Proposition 1. A PI copula of the form (15) has the subset PI property, if for any $I \subseteq G$ the restriction $\overline{A}^{(I)}$, to the I-coordinate space, of the extension $\overline{A}$ of A, is differentiable with locally bounded gradient in that space.

References

Anderson, S.P., A. de Palma and J.-F. Thisse (1992), *Discrete Choice Theory of Product Differentiation*, The MIT Press, Cambridge, MA.

Ben-Akiva, M. and S.R. Lerman (1985), *Discrete Choice Analysis*, The MIT Press, Cambridge, MA.

Joe, H. (1997), Multivariate Models and Dependence Concepts, Chapman and Hall

Keane, M., P. Todd and K.I. Wolpin (2010), The Structural Estimation of Behavioral Models: Discrete Choice Dynamic Programming Methods and Applications, *Handbook of Labor Economics* , eds. Card and Ashenfelter, Volume 2, Elsevier, p. 332-461.

Lau, L.J. (1969), Duality and the structure of utility functions, *J. Ec. Theory* **1**, 374-396.

Li, B. (2011), "The multinomial logit model revisited: A semi-parametric approach in discrete choice analysis", *Transportation Research* B **45**, 461-473.

Lindberg, P.O. (2012a), Probabilistic choice, random utility models, IIA and Mattsson-Weibull-Li distributions", Working Paper, Dept Transp. Science, KTH, Stockholm

Lindberg, P.O. (2012b), Choice probabilities in random utility models — necessary and sufficient conditions for a much used formula", Working Paper, Dept. of Transport Science, KTH Royal Institute of Technology, Stockholm.

Lindberg, P.O., E.A. Eriksson and L.-G. Mattsson (2002), Homothetic Functions Revisited, *Ec. Theory.* **19** 417-427.

Lindberg, P.O., E.A. Eriksson and L-G. Mattsson (1995), Invariance of Achieved Utility in Random Utility Models, *Env. Plan.* **27** 121-142.

Luce, R.D. (1959), *Individual Choice Behavior*, Wiley.

Luce, R.D. and P. Suppes (1965), "Preferences, utility, and subjective probability", in R. Luce, R. Bush, and E. Galanter (eds.), *Handbook of Mathematical Psychology*, Wiley, 249-410.

McFadden, D. (1978), "Modelling the choice of residential location", in A. Karlqvist et al. (eds.), *Spatial Interaction Theory and Planning Models*. North-Holland, 75-96.

Mattsson L.-G. and J.W. Weibull (2010), Extreme-Value Distributions of Wide Applicability, Working Paper, Jan 2010, Dept. Transp. Science, KTH Royal Institute of Technology, Stockholm

Mattsson L.-G., J.W. Weibull and P.O. Lindberg (2011), Extreme values, invariance and choice probabilities", Working Paper, Dept Transp. Science, KTH Royal Institute of Technology, Stockholm

Nelsen, R.B. (2006), *An Introduction to Copulas*, 2nd ed., Springer Verlag.

Shephard, R.W. (1953), *Cost and production functions*, Princeton University Press, Princeton.