INTERTEMPORAL DISCRETE CHOICE, RANDOM TASTES
AND FUNCTIONAL FORM

by

John K. Dagsvik

ABSTRACT

An important problem in the analysis of intertemporal choice processes is to separate the effect of unobserved temporal persistent variables from the influence on preferences from past choice behavior (state dependence).

The present paper discusses a behavioral Axiom in the presence of random preferences relative to a discrete alternative set and demonstrates that this Axiom yields joint utility processes that belong to the class of multivariate extremal processes. Specifically, the Axiom states that if there is no effect from past choice behavior on current preferences then the distribution of the current indirect utility conditional on past choice history is independent of the past choice history. When utilities are extremal processes Dagsvik (1988) demonstrated that the corresponding choice process is Markovian with transition probabilities that have a simple structure.

Key words: Intertemporal discrete choice, habit persistence, structural state dependence, Markovian choice processes, extremal processes.

I thank Rolf Aaberge for his valuable suggestions and Anne Skoglund for excellent word processing.
1. Introduction

The purpose of the paper is to provide a theoretical justification for the structure of intertemporal utility processes under pure taste persistence, i.e., when there is no behavioral effect on the individual level from past choice experiences. In the present setting the agents’s planning horizon is one period and the environment is assumed perfectly certain to him. On the other hand, preferences are assumed random to the agent due to his lack of ability to forecast his preferences perfectly. The case with uncertain environment will be treated elsewhere.

Heckman (1981, 1991) and others have discussed the problem of separating the effect from past behavior on current preferences from spurious effects that stem from correlation between current and past choices due to unobservables. It appears that this identification problem cannot be solved without imposing theoretical restrictions in the model. This identification problem is of considerable practical relevance. For example, in analyses of travel demand it is of interest to know when observed correlation between persons choice of a specific transport alternative at different periods of time is a result of preferences being affected by experience with this alternative or simply a result of time persistent unobservables.

In the present paper we postulate a plausible formal characterization of intertemporal random utility models with pure taste persistence and with preferences that are random relative to the agent. Specifically, this characterization is formulated as an axiom as follows: if no structural state dependence is present then the distribution of the current indirect utility function does not depend on the past choice history. In the present paper we apply the Axiom to a subclass of (intertemporal) random utility models, namely the subclass generated from max-stable processes. Recall that max-stable processes are characterized by having finite-dimensional distributions of the multivariate extreme value type. There are two reasons for studying the subclass. First, it turns out to be convenient from a mathematical point of view. Second, Dagsvik (1991) has demonstrated that the subclass of random utility models generated from max-stable processes is dense in the class of random utility models, in the
sense that the corresponding choice probabilities can be approximated arbitrarily closely by
the choice probabilities derived from max-stable processes. This means that there is essentially
no loss of generality by restricting the utility functions to the class of max-stable processes.
Furthermore, by postulating the axiom described above we get the surprising and important
result that given the class of max-stable processes the utility process must be a (multivariate)
extremal process. In Dagsvik (1983) and (1988) it was proved that the choice model generated
from extremal utility processes has the Markov property with a particular simple structure of
the transition probabilities and transition intensities. Consequently, this framework is
convenient in the context of empirical applications.

The organization of the paper is as follows: In Section 2 a random utility framework
is introduced. In Section 3 we discuss the class of max-stable and extremal utility processes
and the implications from the axioms.

2. Preferences

The individual decision-maker (agent) is supposed to have random preferences in the
sense that they depend on tasteshifters which future realizations are uncertain to the agent.
This notion of random preferences is different from the traditional convention in economics
where stochastic utilities usually are unobservables that are assumed perfectly foreseeable
from the agent’s viewpoint. In the psychological literature however, there is a long tradition
dating back to Thurstone (1927) in which utilities are modeled as random. The reason for this
is of course that individuals have been found to behave inconsistently in laboratory choice
experiments in the sense that a given individual makes different choices under identical
experimental conditions. One explanation for this is that the agent’s psychological state of
mind fluctuates from one moment to the next so as to induce unpredictable shifts in his tastes.
Alternatively, the agent is viewed as having difficulties with evaluating the rank order of the
alternatives, cf. Simon (1988). Thus, at any given point in time neither the agent nor the
observing econometrician is able to predict future tasteshifters. However, tasteshifters realized
in the past are known to the agent but unobserved to the econometrician.

We shall only consider the discrete choice case. The continuous choice case is left for future work.

Let $S$ be a finite set of $n$ alternatives, $a_1, a_2, \ldots, a_n$, and let $\mathcal{B}$ be the index set that corresponds to the collection of all non-empty subsets from $S$. For simplicity we assume that the economic budget constraint holds at each moment in time i.e., no savings and borrowing are allowed. To each alternative, $a_i$, there is associated a stochastic process \( \{U_j(t)\} \), where $U_j(t)$ is the agent's (conditional indirect) utility of $a_j$ given the information and choice history at time $t$. There are no transaction costs and the agent therefore chooses $a_j$ at time $t$ if $U_j(t)$ is the highest utility at $t$. Here age (time) is continuous. Let \( \{J(t)\} \) denote the choice process, i.e.,

$$J(t) = j \quad \text{if} \quad U_j > \max_{k \in B} U_k(t)$$

where $B \in \mathcal{B}$ is the choice set. We shall assume that the choice set $B$ is kept constant over time. We shall henceforth, for notational convenience, suppress $B$ in the notation. Let

$$U(t) = (U_1(t), U_2(t), \ldots, U_n(t))$$

and let $F(t; u(t))$, $u(t) \in \mathbb{R}^n$, be the $n$-dimensional distribution function of $U$ at $t_r = (t_1, t_2, \ldots, t_r)$ where $t_1 < t_2 < \ldots < t_r$, i.e.,

$$F(t_r; u(t)) = P(\bigcap_{i=1}^r \{U(t_i) \leq u(t_i)\}).$$

We assume that \( \{U(t)\} \) is separable and continuous in probability. Moreover, we assume that the mapping $u(t_1) \rightarrow F(t; u(t))$ is continuous. This implies that there are no ties, that is.
\[ P(U_i(t) = U_j(t)) = 0. \]

The class of choice models generated from \( \{U(t)\} \) will be called the class of intertemporal random utility models (IRUM).

When the finite dimensional distributions have been specified it is in principle possible to derive joint choice probabilities for a sequence of choices. In practice however, it turns out to be rather difficult to find stochastic processes that are good candidates for utility processes in the sense that they imply tractable expressions for the choice probabilities in the intertemporal context. More importantly, the class of intertemporal random utility models is quite large and it is thus of substantial interest to restrict this class on the basis of theoretical grounds.

One important theoretical problem in this context is to characterize the preferences under different assumptions about the effect from past choice experience.

The main purpose of the present work is to characterize preferences and the choice probabilities when there are no effects from past experiences on future preferences nor on future choice opportunities. To this end we start with the following definition:

**Definition:**

*By pure-taste-persistent preferences (PTPP) we mean that there are no effects on the agent's preferences from previous choices.*

Thus PTPP means that preferences are exogenous relative to the choice process. Heckman (1981) calls PTPP "habit persistence". We prefer however the notion PTPP since habit persistence may yield association to dependence on past choice experience.

We now introduce a fundamental assumption stated as Axiom 1 below.
Axiom 1

The indirect utility, \( \max_k U_k(t) \), is stochastically independent of \( \{J(s), \forall s < t\} \).

Axiom 1 states that the distribution of the indirect utility at time \( t \) across series of observationally identical choice experiments, does not depend on the choice history prior to \( t \). Clearly, this Axiom follows from PTPP. However, PTPP is not necessarily implied by the Axiom. In fact, PTPP would imply that \( U_j(t) \), for each \( j \), is independent of \( \{J(s), \forall s < t\} \).

In order to clarify the interpretation consider the case in which the preferences also depend on a process, \( \eta = \{\eta_t, t > 0\} \), that is random to the observer but perfectly foreseeable to the agent. Since the conventional assumption is that preferences are deterministic to the agent this case thus represents a generalization of the traditional setting. Then the obvious modification of Axiom 1 to be consistent with PTPP is

\[
P(\max_k U_k(t) \leq y | J(s), s < t, \eta) = P(\max_k U_k(t) \leq y | \eta).
\]

(2.1)

If we also require that

\[
P(\max_k U_k(t) \leq y | J(s), s < t) = P(\max_k U_k(t) \leq y)
\]

(2.2)

then the process \( \eta \) must be stochastically independent of \( \{J(s), s < t\} \). The difference in interpretation between (2.2) and (2.1) is the following: While (2.1) follows from PTPP (on the individual level) (2.2) corresponds to an aggregate analogue - namely that PTPP holds on average.

In Sections 3 and 4 we shall investigate the implications from Axiom 1.

3. The class of intertemporal generalized extreme value models (IGEV)

The IGEV is generated from utility processes that are max-stable. A max-stable process has finite dimensional distributions of the multivariate extreme value type. This means
that maximum of independent copies of a max-stable process is max-stable (cf. de Haan, 1984). As is wellknown, there are three types within this class. We shall here consider type III which has finite-dimensional distributions (normalized) that are characterized by

\[
\log F(t, u(t, )) = e^{-z \log F(t, u(t, )) - z1}, \ z \in \mathbb{R}
\]

where \( 1 = (1, 1, \ldots, 1) \). In particular, the one-dimensional marginal distribution has the form

\[
F_i(t; u_i) = \exp(-e^{y_i u_i}).
\]

In Dagsvik (1991) it is proved that under suitable conditions the class of IGEV models is dense in the class of IRUM. By this it is meant that IRUM choice probabilities for a specific sequence of alternatives can be approximated arbitrarily closely to the corresponding choice probabilities of an IGEV model. The implication of this result is that there is no loss of generality by restricting the class of IRUM to the class of IGEV. Thus in the following we shall assume that the utility function is a multivariate max-stable process which is continuous in probability.

A very important subclass of the class of max-stable processes is called extremal processes. Multivariate extremal processes have been characterized by de Haan and Resnick (1977) and Dagsvik (1988). Following Dagsvik (1988) let \( \{H_t, Q_0\} \) be a family of multivariate extreme value distribution functions that satisfies \( \sum_h Q_0 = 1 \). Suppose also that \( H_i(w)/H_s(w) \), for \( t > s \), is a distribution function in \( \mathbb{R}^n \). The multivariate extremal process has the same finite-dimensional distribution as \( \{Y(t), t \geq 0\} \) defined by

\[
Y(t) = \max(Y(s), W(s,t)), \ s < t, \ Y(0) = -\infty
\]

where maximum is taken componentwise, \( W(s,t) \) and \( W(s',t') \) are independent when \( (s,t) \cap (s',t') = \emptyset \) and
\[ P(W(s,t) \leq w) = \frac{H_i(w)}{H_i(\bar{w})}. \] (3.3)

Also \( Y(s) \) and \( W(s,t) \) are independent. It can be demonstrated that a multivariate extremal process is a pure jump Markov process (cf. de Haan and Resnick, 1977).

In Appendix 2 we summarize some of the properties of multivariate extremal processes. The class of multivariate extremal processes turns out to be of particular importance in choice theory as the next result indicates.

**Theorem 1**

Assume that the choice model belongs to the IGEV class with utility process that is continuous in probability. Then Axiom 1 implies that the utility function is a multivariate extremal process.

The proof of Theorem 1 is given in the appendix.

Theorem 1 provides the necessary theoretical foundation for postulating utility processes that are of the extremal type, at least as a point of departure. Theorem 1 has some very important implications which are summarized in Theorem 2 and Corollary 1.

**Theorem 2**

Suppose that the utility function is a multivariate extremal process (type III) with c.d.f. \( H_i(y) \) at \( t \). Then the choice process \( \{J(t)\} \) is a Markov chain with marginal - and transition probabilities given by

\[ P_j(t) \equiv P(J(t) = j) = \frac{\partial_i G_i(\theta)}{G_i(\theta)}, \] (3.4)
Theorem 2 has been proved in Dagsvik (1983) in the case where utilities are independent across alternatives. The more general case with interdependent utilities is proved in Dagsvik (1988). Resnick and Roy (1990) give a proof that does not require that the partial derivatives of $H_i(y)$ exist.

From Theorem 2 it is easy to obtain the corresponding transition intensities. Recall that the transition intensites for $\{J(t)\}$ are defined by

$$Q_{ij}(s,t) = \lim_{\Delta t \to 0} \frac{Q_{ij}(s,t + \Delta t)}{\Delta t}, \text{ for } i \neq j$$

and

$$Q_{ii}(s,t) = \lim_{\Delta t \to 0} \left( \frac{Q_{ii}(s,t + \Delta t) - 1}{\Delta t} \right).$$

According to Dagsvik (1988) we have the following:

**Corollary 1**

Suppose $G_i(y)$ is differentiable with respect to $t$. Let $g_i(y) \equiv \partial G_i(y) / \partial t$. Then

$$\gamma_{ij}(t) = \frac{-\partial g_j(0)}{G_i(0)}, \quad \text{for } i \neq j, \text{ and}$$

$$\gamma_{ii}(t) = \frac{Q_{ii}(t,t + \Delta t) - 1}{\Delta t}.$$
\[ \gamma_i(t) = -\sum_{k\neq i} \gamma_{ik}(t) = \frac{\partial g_i(0)-g_i(0)}{G_i(0)}. \quad (3.8) \]

4. Interpretation of results

From Theorem 1 and Corollary 1 we can draw a number of important conclusions. First, note that (3.4) is the well-known formula for the choice probability in the GEV class (Generalized extreme value models), cf. McFadden (1981). The interpretation of \( G_i(0) \) is as

\[ E(\max_k U_k(t)) = \log G_i(0) + 0.5772 \ldots \quad (4.1) \]

Similar to Dagsvik (1983) it can be proved that

\[ \text{corr}\left\{ \max_k U_k(s), \max_k U_k(t) \right\} = \rho \left( \frac{G_i(0)}{G_i(0)} \right), \quad s \leq t \quad (4.2) \]

where \( \rho: [0,1] \rightarrow [0,1] \) is an increasing function with \( \rho(0)=0 \) and \( \rho(1)=1 \). By combining (3.4) and (3.5) we get

\[ Q_{ij}(s,t) = P_j(t) - \frac{G_i(0)P_j(s)}{G_i(0)}, \quad s \leq t, \quad i \neq j \quad (4.3) \]

and

\[ Q_{ii}(s,t) = P_i(t) + \frac{G_i(0)}{G_i(0)} (1-P_i(s)) \quad (4.4) \]

where we emphasis that, apart from a monoton mapping \( G_i(0)/G_i(0) \) is the autocorrelation function of the indirect utility function. In other words, in the absence of structural state dependence there is a simple relationship between the transition probabilities, the marginal probabilities (3.4) and the autocorrelation function of the indirect utility function.
The transition intensities (3.7) and (3.8) can be given a particular interpretation. Let

\[ p_j(t) = -\frac{\partial g_i(0)}{g_i(0)}. \]  

(4.5)

Recall that \( \{\max_k U_k(t)\} \) is a pure jump Markov process. The interpretation of (4.5) is that given that a jump of the process \( \{\max_k U_k(t)\} \) occurs at time \( t \), \( p_j(t) \) is the probability that the highest utility is attained in state \( j \). To see this note that by Theorem A1 (iv) in Appendix 2

\[ 1 - \frac{g_i(y)}{g_i(x1)} \quad \text{for} \quad y > x1, \ x \in \mathbb{R}, \ y \in \mathbb{R}^n \]

is the c.d.f. of the utility process given that a jump occurs at \( t \) and given that \( \max_k U_k(t) = x \).

Note also that \( g_i(y) \) has the property

\[ g_i(y) = e^{-\gamma}g_i(y-x1) \]  

(4.6)

which follows immediately from (3.1). The probability that \( J(t) = j \), given that a jump occurs at \( t \) and given \( \max_k U_k(t) = x \), is therefore easily demonstrated to be

\[ -\int_x^{\infty} g_i(y1)dy \quad \text{for} \quad y \in \mathbb{R} \]

\[ \frac{\partial g_i(y1)}{g_i(x1)}, \ y \in \mathbb{R} \]

which by (4.6) reduces to

\[ -\frac{\partial g_i(0)}{g_i(0)} \int_x^{\infty} e^{-\gamma y}dy = p_j(t). \]

From (3.7) we thus obtain that
\[ \gamma'_j(t) = p_j(t) \frac{g_i(0)}{G_i(0)} = p_j(t) \partial_t E(\max_k U_k(t))/\partial t, \quad \text{for } i \neq j. \] (4.7)

The last equality follows from the wellknown property that

\[ E(\max_k U_k(t)) = \log G_i(0) + 0.5772. \]

The second equality of (4.7) tells us that the transition intensity is the product of the rate of change in the mean indirect utility and the conditional probability that the highest utility is attained at alternative j given that a jump of the indirect utility occurs at t.

The conditional probability, \( \pi_{ij}(t) \), of moving to state j given that state i is left follows from

\[ \gamma'_j(t) = \pi_{ij}(t) \sum_{i \neq j} \gamma'_i(t) = -\pi_{ij}(t) \gamma'_i(t). \]

Hence (4.7) implies

\[ \pi_{ij}(t) = \frac{p_j(t)}{1 - p_i(t)}. \] (4.8)

The results discussed above are related to the particular utility function representation. However, we can use Theorem 1 and Theorem 2 to obtain a characterization result of the choice model that is independent of the particular utility structure provided the utility process \( \{U(t), t > 0\} \) satisfies the condition

\[ E(\max(1, \exp(a (\sup_{s \in K} (\max_k U_k(s)))))) < \infty \] (4.9)

for some \( a > 0 \) and any Borel set \( K \subset (0, \infty) \) that is compact.
Theorem 3

Assume that the choice model is a random utility model where the utility process satisfies (4.9) and the first order partial derivatives of the corresponding finite-dimensional distributions exist. Assume furthermore that Axiom 1 holds. Then the choice process \( \{J(t), t>0\} \) is a Markov chain and the transition probabilities, \( \{\tilde{Q}(s,t)\} \), from state (alternative) \( i \) to state \( j \) have the form

\[
\tilde{Q}(s,t) = P(J(t) = j) - K(s,t)P(J(s) = j) \quad \text{for} \quad j \neq i,
\]

(4.10)

and

\[
\tilde{Q}(s,t) = P(J(t) = i) + (1 - P(J(s) = i))K(s,t)
\]

(4.11)

for some suitable positive function, \( K(s,t) \).

Proof:

Assume first that the random utility model belongs to the IGEV class. Then by Theorem 1 and 2 \( \{J(t), t>0\} \) is a Markov chain with transition probabilities that have the structure given in (4.3) and (4.4). Dagvski (1992) has proved that under condition (4.9) the class of IGEV models is dense in the class of random utility models. This means that for any \( \delta > 0 \) it is possible to find an IGEV model with transition probabilities \( \{Q(s,t)\} \) such that

\[
|P(J(t) = j| J(t) = i, J(t_r), \forall r < n) - Q(s,t)| < \delta
\]

(4.12)

holds for any \( t_1 < t_2 < \ldots < t_n < t \).

Suppose now that the Theorem is not true, i.e., for some time epochs \( t_1, t_2, \ldots, t_n, t \) and some choice sequence

\[
|P(J(t) = j| J(t) = i, J(t_r), \forall r < n) - \tilde{Q}(s,t)| > \delta
\]
for any choice of \( K(s,t) \). But since \( \tilde{Q}_t(s,t) \) has the same structure as \( Q_t(t,s) \) we may choose \( K(s,t) \) such that \( \tilde{Q}_t(s,t) = Q_t(s,t) \) which yields a contradiction of (4.12). Thus \( \{ J(t), t > 0 \} \) is a Markov chain and the transition probabilities have the structure stated in the Theorem.

Q.E.D.

The result in Theorem 3 is of particular interest - since it provides a nonparametric test for the hypothesis that no structural state dependence effects are present.

Recall that in the present paper we have assumed that the choice set is constant over time\(^3\) and that there are no transaction costs.

From Corollary 1 it follows that the hazard function, \( h_i(t) \), of state \( i \) is given by

\[
    h_i(t) = -\gamma_i(t) = \frac{g_i(0) - \partial g_i(0)}{G_i(0)}.
\]  

Let \( T_i(s) \) be the length of time in state \( i \) given that state \( i \) was entered at time \( s \). Then it follows immediately from (4.13) that the survivor function of the choice process \( \{ J(t) \} \) is given by

\[
P(T_i(s) > t - s) = \exp \left(- \int_s^t h_i(u) du \right) = \frac{G_i(0)}{G_i(0)} \exp \left(- \int_s^t \frac{\partial g_i(0)}{G_i(0)} du \right).
\]
5. Special cases

In the case where the utilities are independently distributed across alternatives we obtain

\[ G_i(y) = \exp(\beta(t)) \sum_k \exp(v_k(t) - y_k) \]  

(5.1)

and

\[ p_j(t) = \frac{\exp(v_j(t))}{\sum_k \exp(v_k(t))} \]  

(5.2)

where \( \beta(t) \) is some function of \( t \) and where \( v_j(t) \) has the interpretation

\[ EU_j(t) = v_j(t) + \beta(t). \]  

(5.3)

From Corollary 1 we get that the transition intensities in this case reduce to

\[ \gamma_{ij}(t) = p_i(t)(v_j(t) + \beta'(t)) = \frac{(v_j'(t) + \beta'(t))\exp(v_j(t))}{\sum_k \exp(v_k(t))} \]  

(5.4)

for \( i \neq j \), provided \( v_j(t) \) and \( \beta(t) \) are differentiable. We see from (5.4) that we must have \( v_j'(t) + \beta'(t) > 0 \) for all \( j \). By (4.13) the corresponding hazard function reduces to

\[ h_i(t) = \sum_k p_k(t)v_k'(t) + \beta'(t) - p_i(t)(v_i'(t) + \beta'(t)). \]  

(5.5)

Consider next the particular case with independent utilities and \( v_j(t) = y_j \). Then (5.5) reduces to

\[ h_i(t) = \beta'(t)(1 - p_i) \]  

(5.6)

and the survivor function (4.14) becomes
\[ P(T_i(s) > t - s) = \exp(-(1 - P_i)(\beta(t) - \beta(s))) \]  
\[ P(T_i(s) > t - s) = \exp(-(1 - P_i)(\theta_0 s + \theta) \)  

where \( P_i \) is given by (5.2) when the mean utilities \( v_k(t) \) are substituted by \( v_k, k=1,2,...,n \).

If in addition \( \beta(t) \) is linear in \( t; \beta(t)=t\theta \), then \( \{J(t)\} \) becomes a time homogenous Markov chain with exponentially distributed holding times i.e.,

\[ P(T_i(s) > t - s) = \exp(-(t - s)(1 - P_i)\theta) \]  

and transition intensities

\[ \gamma_{ij} = \theta P_j \quad \text{for } i \neq j \]  
\[ h_i = -\gamma_{ii} = \theta(1 - P_i). \]  

There is an important observation to be made here. Recall that the well-known specification of the transition intensities due to Cox (see Andersen et al., 1991) in the multistate case can be written

\[ \gamma_0(t) = \lambda_0(t) \exp(f(Z_i, Z_j; \alpha)) \]  

where \( f(\cdot) \) is some specified function and \( Z_i \) is an individual specific time invariant vector of state specific covariates that characterize state \( i \) and \( \alpha \) is a vector of parameters. \( \lambda_0(t) \) is called the baseline hazard. Let us now compare the structure (5.11) with the result of Corollary 1 in the time homogenous case when the utilities are independent across alternatives and

\[ v_i = f(Z_i, \alpha). \]

Then we get
\[ \gamma_0(t) = \beta'(t) \frac{\beta'(t) \exp \left( f(Z_i, \alpha) \right)}{\sum_k \exp \left( f(Z_k, \alpha) \right)} \quad i \neq j. \quad (5.13) \]

We realize that (5.13) is essentially different from (5.11) in that it depends on all the covariates in a particular way while (5.11) only depends on the covariates related to state i and j. Therefore, the standard proportional hazard specification (5.11), which is often applied in duration analysis, is inconsistent with a random utility formulation.
Footnotes

1) Recall that a process \( \{Y(t), t>0\} \) being continuous in probability means that \( \exists \delta > 0 \) such that for any \( \eta_1, \eta_2 > 0 \)

\[
P(\|Y(t) - Y(s)\| > \eta_1) < \eta_2
\]

whenever \( |t-s| < \delta \), where \( \| \cdot \| \) is the standard Euclidian metric.

2) The function \( \rho(\cdot) \) has the form

\[
\rho(y) = -\frac{6}{\pi^2} \int_0^y \frac{\log{x}}{1-x} \, dx
\]


3) It is possible to extend the results of this paper to the case with non-decreasing choice sets. (See Dagsvik, 1983).
Lemma 1

Let $F(x,y)$ be a bivariate (type III) extreme value distribution. Then $-\log F(-x,-y)$ is convex. If $F(x,y)$ is continuous the left and right derivatives, $\partial F_2(x,y)/\partial x$ and $\partial F_2(x,y)/\partial y$, exist and are non-decreasing.

Proof:

Let $L(x,y) = -\log F(-x,-y)$. Since $F$ is a c.d.f. it follows that $L$ is non-decreasing. Moreover, since $F$ is a bivariate extreme value distribution it follows by Proposition 5.11, p. 272 in Resnick (1987) that there exists a finite measure $\mu$ on

$$\Delta = \{z \in \mathbb{R}^2 : z_1^2 + z_2^2 = 1\}$$

such that

$$L(x,y) = \int \max(z_1 e^x, z_2 e^y) \mu(dx, dy).$$

Since $z_1 e^x$ and $z_2 e^y$ are convex functions it follows that $L(x,y)$ is convex. Since $L(x,y)$ is convex the left and right derivatives of $F(x,y)$ exist. (See for example Kawata, Theorem 1.11.1 p. 27.)

Q.E.D.

Proof of Theorem 1

Resnick and Roy (1990) have demonstrated that there is no loss of generality by assuming the components of the utility process are independent. Although their discussion only regards multivariate extremal processes it is clear that their argument goes through also in the general max-stable case. Thus we assume that the components $U_j(t)$, $j=1,2,...,n$, of $U(t)$ are independent. Also we assume $n=2$ since the general case is completely analogous.
Consider the case of choices at two moments in time, s and t. Let \( F(s,t; x_1,x_2,y_1,y_2) \) be the corresponding c.d.f. of \((U(s), U(t))\). By the independence assumption

\[
F(s,t; x_1,x_2,y_1,y_2) = F_1(s,t; x_1,y_1)F_2(s,t; x_2,y_2)
\]  

(A.1)

where \( F_j \) is the c.d.f. of \((U_j(s), U_j(t))\). Note that since \( \{U_j(t), t \geq 0\} \) is assumed continuous in probability it follows that \( F_j(s,t; x,y) \) is continuous in \((x,y)\).

Let

\[
G_i(x,y) = -\log F_i(s,t; x,y).
\]

The probability distribution, \( M(s,t; j_1, j_2, z) \), of \( \{J(s), J(t), \max_k U_k(t)\} \) is given by

\[
M(s,t; 2,1,z) = \mathbb{P}(J(s)=2, J(t)=1, \max_k U_k(t) \leq z) = \int F(s,t; x, dx, dy, y). 
\]

(A.2)

and similarly for other values of \( j_1 \) and \( j_2 \).

By Lemma 1 the first order left and right derivatives of \( G_i(x,y) \) exist. From now on we shall use the notion derivative of \( G_i(x,y) \) meaning the respective (first order) right derivatives.

Since \( F \) by assumption is a multivariate extreme value distribution it follows from (3.1) that for \( z \in \mathbb{R} \)

\[
G_i(x,y) = e^{-z}G_i(x-z,y-z).
\]

Hence

\[
F(s,t; x, dx, dy, y) = [\exp(-e^{-z}(G_1(x-y,0)+G_2(x-y,0)))]
\cdot e^{-2z}\partial_2G_1(x-y,0)\partial_1G_2(x-y,0)dx dy 
\]

(A.3)

where \( \partial_j \) denotes the partial derivative with respect to component \( j \). Let
From the relationship

$$G_i(x,y) = e^{-\gamma} G_i(x-y,0) = e^{-\gamma} h_i(y-x)$$

it follows that

$$\partial^2 G_i(-x,0) = -h_i(x) + h'_i(x). \quad (A.4)$$

By Lemma 1 $h_i(x)$ is convex and has derivatives that are non-decreasing. From (A.2), (A.3) and (A.4) it follow after the change of variable $x=-u+y$ that

$$M(s,t; 2,1,dy) = e^{-2\gamma dy} \int_k [\exp(-\gamma(h_i(u)+h_2(u)))] h'_i(u)(h_2(u)-h_2'(u))du. \quad (A.5)$$

Suppose now that $x = r \geq -\infty$ is the largest point at which $h_i'(x) + h_2'(x) = 0$. Then since $h_i'(x)$ is nondecreasing it must be true that $h_i'(x) = h_2'(x) = 0$ for $x \leq r$. As a consequence the mapping $\psi : R_+ \to [r,\infty)$ defined by

$$z = h_i(\psi(z)) + h_2(\psi(z)) - h_i(0) - h_2(0) \quad (A.6)$$

exists, is invertible and has (right) derivative everywhere on $R_+$. By change of variable

$$u \to \psi^{-1}(u) = z$$

(A.5) takes the form

$$M(s,t; 2,1,dy) = e^{-2\gamma dy} \int \exp(-\gamma z) \gamma(z)dz \quad (A.7)$$
where
\[ q = h_1(-\infty) + h_2(-\infty) - h_1(0) - h_2(0) = h_1(t) + h_2(t) - h_1(0) - h_2(0), \]

\[ \gamma(z) = h_1'(\psi(z))\psi'(z)(h_2(\psi(z)) - h_2'(\psi(z))) \]  \hspace{1cm} (A.8)

and
\[ b = h_1(0) + h_2(0). \]

According to Axiom 1 we must have
\[ M(s,t; ij, dy) = M(s,t; ij, \infty)P(\max_k U_k(t) = dy). \]  \hspace{1cm} (A.9)

The (marginal) distribution of \( \max_k U_k(t) \) for \( i, j = 1,2, \) is easily demonstrated to be extreme value as below

\[ P(\max_k U_k(t) \leq y) = \exp(-e^{-\gamma b}). \]  \hspace{1cm} (A.10)

With \( \theta = \exp(-y) \) we thus obtain from (A.5), (A.6) and (A.7)

\[ \theta e^{-\beta} \int_{q}^{\infty} (\exp(-\theta z))\gamma(z)dz = M(s,t; 2,1,\infty)be^{-\beta}. \]  \hspace{1cm} (A.11)

Note that (A.11) implies that the Laplace transform of \( \gamma(z) \) has the form \( c/\theta \) where \( c \) is a constant. But this implies that \( \gamma(z) = 0 \) for \( z < q \) and

\[ \gamma(z) = M(s,t; 2,1,\infty)b, \quad z \geq q. \]  \hspace{1cm} (A.12)

From the definition of \( \psi(z) \) we get

\[ 1 = (h_1'(\psi(z)) + h_2'(\psi(z))\psi'(z)). \]  \hspace{1cm} (A.13)
Hence (A.8), (A.12) and (A.13) with \( u = \psi(z) \) yield

\[
h_1'(u)(h_2(u) - h_2'(u)) = h'(u)C_{21}, \text{ for } u > r,
\]

where \( h(u) = h_1(u) + h_2(u) \) and \( C_{ij} = M(s, t; i, j, \infty)b \). Similarly we get

\[
h_2'(u)(h_1(u) - h_1'(u)) = h'(u)C_{12}, \text{ for } u > r.
\]

By subtracting (A.14) from (A.15) we get

\[
h'(u)h_1(u) - h'_1(u)h(u) = h'(u)(C_{12} - C_{21})
\]

which, when dividing by \( h(u)^2 \) becomes equal to

\[
\frac{h_1'(u)h(u) - h_1(u)h'(u)}{h(u)^2} = \frac{h'(u)(C_{21} - C_{12})}{h(u)^2}.
\]

Next, integrating both sides of (A.16) yields

\[
\frac{h_1(u)}{h(u)} = \frac{C_{12} - C_{21}}{h(u)} + d, \text{ for } u > r,
\]

where \( d \) is a constant. Hence we obtain

\[
h_1(u) = C_{12} - C_{21} + h(u)d, \text{ for } u > r.
\]

By inserting (A.17) into (A.14) we get

\[
h'(u)(h_2(u) - h_2'(u))d = h'(u)C_{21}, \text{ for } u > r,
\]

which is equivalent to
\begin{align}
\tag{A.18a}
h_u(u) - h'_u(u) &= C_{21}/d.
\end{align}

Similarly
\begin{align}
\tag{A.18b}
h_i(u) - h'_i(u) &= C_{12}/d, \text{ for } u > r.
\end{align}

Eq. (A.18) is a first order differential equation which has a solution of the form
\begin{align}
\tag{A.19a}
h_j(u) &= \alpha_j + \beta_j e^u, \text{ for } u > r, \ j = 1, 2.
\end{align}

Since \( h'_j(u) = 0 \) for \( u \leq r \) and \( h_j(u) \) is continuous we get from (A.19a) that
\begin{align}
\tag{A.19b}
h_j(u) &= \alpha_j + \beta_j e^r, \text{ for } u \leq r.
\end{align}

As a consequence
\begin{align}
G_j(x, y) &= e^{-y}h_j(y-x) = \alpha_j e^{-y} + \beta_j \exp(-\min(x, y-r)).
\tag{A.20}
\end{align}

From (A.20) we obtain that for \( s < t \)
\begin{align}
P(U_j(t) \leq y | U_j(s) = x) &= 0 \text{ when } y < x + r \tag{A.21}
\end{align}

and
\begin{align}
P(U_j(t) \leq y | U_j(s) = x) &= P(U_j(t) \leq y) \text{ when } y > x + r. \tag{A.22}
\end{align}

Eq. (A.21) means that \( \{U_i(t)\} \) is non-decreasing with probability one. Eq. (A.22) means that conditional on \( U_i(t) > U_i(s) \) then \( U_i(t) \) is stochastically independent of \( U_i(s) \). But then we must have that \( \{U_i(t)\} \) is equivalent to the utility process defined by
\begin{align}
U_i(t) &= \max(U_i(s), W_i(s, t)) + r \tag{A.23}
\end{align}

where \( W_i(s, t) \) is extreme value distributed and independent of \( U_i(s) \). Since \( U_1(t) - U_2(t) \) is
independent of r for any t we may without loss of generality choose r=0. But then (A.23) defines the extremal process as defined by Tiago de Oliveira and others, (cf. Dagsvik, 1983, 1988) which was to be proved.

So far we have proved that conditional on a particular choice history, at two points in time, Axiom 1 implies utilities that are extremal processes. We have not yet demonstrated that the class of choice models with extremal utility processes fullfills the requirement of Axiom 1 when we condition on \( \{J(t), \forall t \leq t\} \). Fortunately, however, this has been proved by Resnick and Roy (1990), p.p. 321.

Q.E.D.

APPENDIX 2

Theorem A1

Suppose \( H_i(w) \) is differentiable in \( t \). Then the multivariate extremal process \( \{Y(t)\} \) has the following properties

(i) It is continuous in probability.

(ii) With probability one it has only a finite number of jumps in a finite time interval.

(iii) The transition probability function is given by

\[
P(Y(t) \leq y | Y(s) = w) = \begin{cases} 
H_i(y)/H_i(y), & y \geq w, \; s < t \\
0, & \text{otherwise}
\end{cases}
\]  

(A.29)

(iv) Given that a jump occurs the process jumps from \( x \) into \( [y, \infty) \) with probability
\[ \Pi_r(w, [y, \infty)) = \begin{cases} \frac{g_r(y)}{g_r(w)} & \text{if } y > w \\ 0 & \text{if } y \leq w \end{cases} \quad (A.30) \]

where

\[ g_r(y) = \frac{-\partial \log H_r(y)}{\partial t}. \quad (A.31) \]

**Proof:**

(i) This result is demonstrated by extending the proof by Resnick (1987) p. 182 to the multivariate non-homogeneous case.

(ii) This is Theorem 1 in Dagsvik (1988).

(iii) Dagsvik (1988), p. 33 gives this result.

It thus only remains to prove (iv). From (iii) it follows that the intensity of a jump out of state \( w \) is given by

\[ \lim_{t \to s}(P(Y(t) > w | Y(s) = w)/(t-s)) = g_r(w). \]

Recall that \( \Pi_r(w, [y, \infty)) \) is the probability that \( \{Y(t)\} \) jumps from \( w \) into \([y, \infty)\) given that a jump occurs. Since \( \{Y(t)\} \) is a Markov process we have

\[ \lim_{t \to s}(P(Y(t) > y | Y(s) = w)/(t-s)) = g_r(w)\Pi_r(w, [y, \infty)). \]

But by (iii) we get

\[ \lim_{t \to s}(P(Y(t) > y | Y(s) = w)/(t-s)) = g_r(y) \]

for \( y \geq w \) and zero otherwise. Thus by combining the last two equations yields
\[ \Pi_{i}(w,[y,\infty)) = \frac{g_{i}(y)}{g_{i}(w)} \]

Q.E.D.
REFERENCES


No. 7 P. Frenger (1985): A Directional Shadow Elasticity of Substitution.


No. 28  J.K. Dagsvik (1987): A Modification of Heckman's Two Stage Estimation Procedure that is Applicable when the Budget Set is Convex.


No. 52 R. Aaberge and T. Wennemo (1990): Non-Stationary Inflow and Duration of Unemployment.


No. 56 N.M. Stølen (1990): Is there a NAIRU in Norway?


No. 66 E. Bowits and E. Storm (1991): Will restrictive demand policy improve public sector balance?


No. 70 R. Aaberge, Xiaojie Chen, Jing Li and Xuezeng Li (1992): The structure of economic inequality among households living in urban Sichuan and Liaoning, 1990.


No. 72 B. Bye and E. Holmøy (1992): Dynamic equilibrium adjustments to a terms of trade disturbance

No. 73 O. Aukrust (1992): The Scandinavian contribution to national accounting

No. 74 J. Aasness, E, Eide and T. Skjerpen (1992): A criminometric study using panel data and latent variables (will be issued later)

No. 75 R. Aaberge and Xuezeng Li (1992): The trend in income inequality in urban Sichuan and Liaoning, 1986-1990

No. 76 J.K. Dagsvik and Steinar Strøm (1992): Labor supply with non-convex budget sets, hours restriction and non-pecuniary job-attributes

No. 77 J.K. Dagsvik (1992): Intertemporal discrete choice, random tastes and functional form