Discussion Papers
Statistics Norway
Research department

No. 688
May 2012

Statistics Norway

Statistisk sentralbyrå

John K. Dagsvik

Behavioral multistate duration models

What should they look like?

Discussion Papers No. 688, May 2012
Statistics Norway, Research Department

John K. Dagsvik

Behavioral multistate duration models
What should they look like?

Abstract:

This paper discusses how specification of probabilistic models for multistate duration data generated by individual choices should be justified on a priori theoretical grounds. Preferences are assumed represented by random utilities, where utilities are viewed as random also to the agent himself. First, the paper proposes a characterization of exogenous preferences, (that is, in the special case with no state dependence effects). The main assumption asserts that when preferences are exogenous the current and future indirect utilities are uncorrelated with current and past choices, given unobservables that are perfectly known to the agent. It is demonstrated that under rather weak and general regularity conditions this characterization yields an explicit structure of the utility function as a so-called Extremal stochastic process. Furthermore, from this utility representation it follows that the choice process is a Markov Chain (in continuous- or discrete time), with a particular functional form of the transition probabilities, as explicit functions of the parameters of the utility function and choice set. Subsequently, we show how the model can be extended to allow for structural state dependence effects, and how such state dependence effects can be identified. Moreover, it is discussed how a version of Chamberlain's conditional estimation method applies in the presence of fixed effects. Finally, we discuss two examples of applications.

Keywords: Duration models, Random utility models, Habit persistence, True state dependence, Extremal process, Markov chain

JEL classification: C23, C25, C41, C51, D01

Acknowledgements: This research has been supported by Statistic Norway, the Frisch Centre for Economic Research, Oslo, and the Department of Economics, University of Turin. Part of this research has been carried out while the author was visiting Professor at the Department of Economics, University of Turin (Contract no 08/CVP). I am grateful for comments by Zhiyang Jia, Steinar Strøm and Nils Martin Stølen.

Address: John K. Dagsvik, Statistics Norway, Research Department and the Frisch Centre for Economic Research. E-mail: John.Dagsvik@ssb.no

Discussion Papers comprise research papers intended for international journals or books. A preprint of a Discussion Paper may be longer and more elaborate than a standard journal article, as it may include intermediate calculations and background material etc.

Abstracts with downloadable Discussion Papers
in PDF are available on the Internet:
http://www.ssb.no
http://ideas.repec.org/s/ssb/dispap.html

For printed Discussion Papers contact:
Statistics Norway
Telephone: +47 62 88 55 00
E-mail: Salg-abonnement@ssb.no

ISSN 0809-733X
Print: Statistics Norway

Sammendrag

Denne artikkelen diskuterer hvordan spesifikasjon av probabilistiske modeller for fler-tilstands varighetsdata generert fra individuelle valg kan begrunnes. Preferansene til aktørene antas representerte ved en stokastisk nyttefunksjon som antas stokastisk også for aktøren. Artikkelen starter med å foreslå en karakterisering av eksogene preferanser, dvs. si spesialtilfellet uten strukturell tilstandsavhengighet. Denne karakteriseringen sier at når preferansene er eksogene, så vil den indirekte nytten ved et gitt tidspunkt ikke være korrelert med nåværende og tidligere valg, gitt de uobserverbare variablene som er kjent for aktøren. Det vises at denne antakelsen, i tillegg til svake regularitets-betingelser, impliserer en nyttefunksjon som er en såkalt Extremal stokastisk prosess. Videre medfører dette resultatet at valgmodellen er en Markov kjede (i kontinuerlig- eller diskret tid), med overgangs-sannsynligheter som har en bestemt funksjonsform, som eksplisitt funksjon av parametrene i nytte-funksjonen. Deretter diskuteres det hvordan modellrammen kan utvides til å ta hensyn til strukturell tilstandsavhengighet. Det diskuteres også hvordan en versjon av Chamberlains betingede estimerings-metode kan anvendes når en har fixed effects. Til slutt diskuteres to eksempler på anvendelser.

1. Introduction

The issue of functional form in behavioral relations in the social sciences is a problematic one. In contrast to physics where the theory of the phenomenon under study often yields a complete characterization of the corresponding quantitative structural relations, the theories in social sciences are typically qualitative and imply little guidance as regards explicit mathematical functional form.

This paper discusses how to justify functional form relations of duration models. More precisely, we consider how hazard functions and transition probabilities generated by individual choice behavior in continuous-or discrete time should be modeled.[1] We assume a probabilistic formulation in which preferences are represented by utilities that not only are perceived as random to the researcher, but are also viewed as random to the individual agent. Psychologists have found that preferences often appear to be random to the agent himself, cf. Thurstone (1927) and Tversky (1969). The explanation is that the agent may express some uncertainty in their evaluation of alternatives and consequently assess different values to the same, or seemingly equivalent alternative, at different points in time. This includes updating of preferences due to information that arrives at random points in time (as perceived by the agent).[2]

In general, preferences may be serially correlated due to habit persistence, structural state dependence and unobserved heterogeneity. Unobserved heterogeneity stems from unobservables that may be perfectly known, or partly uncertain, to the agent. See Heckman (1981a) for a lucid discussion within a multivariate probit framework. In several analyses it is of interest to separate the genuine effect on preferences (or choice constraints) due to choice experience on one hand (structural state dependence), from the effect from unobserved heterogeneity and habit persistence on the other. It is known that this identification problem cannot be settled by statistical methods alone. Additional a priori theoretical restrictions are needed, see for example Heckman, (1981a,b, 1991), Jaggia and Trivedi (1994) and Magnac (2000). The theory proposed in this paper implies an explicit functional form characterization of the model that brings us a step further towards resolving the identification of state dependence effects.

We first provide a characterization of the individual's preferences and the implied choice model in the special case with no state dependence, conditional on the information of the agent. That is, we propose a characterization of the utility representation and the corresponding choice model,

[1] For an overview of duration model with particular reference to economics, see van den Berg (2001).

[2] This type of uncertainty is thus different from the lottery setting with uncertain outcomes, because although the utilities may vary from moment to moment in a manner that is not fully predictable to the agent, the decision process consists of selecting the alternative with the largest momentary utility, and, at the moment of choice, the utilities are known to the agent.

conditional on unobservables that are perfectly known to the agent. Sometimes it may be reasonable to represent unobservables by fixed or random effects. It follows from the setting described above that since preferences may contain elements that are random to the agent himself the conditional utility representation given the agents (current) information may be random (to the agent). Moreover, even in the case with no state dependence there may be habit persistence implying that the conditional preferences are serially correlated (given the information of the agent). Subsequently, we extend the characterization to the case that allows for structural state dependence. Let the "reference case" be defined as the special case with no state dependence. In this case we propose an intuitive definition of what one should understand by random preferences being exogenous. Specifically, this definition asserts that in the reference case, the indirect utility (conditional on unobservables known by the agent) at any current or future point in time is independent of current and previously chosen alternatives (states). The underlying intuition is that since the utility function represents the value of the respective states, once maximum utility has been achieved, it should (in the reference case) be irrelevant for level of maximum utility which of the states that yields the maximum. This assumption seems almost inescapable: In fact, if this were not so, not all relevant aspects of the states would be captured by the utility function since knowledge about the actual choice and the choice history would represent relevant information about the level of the utility of the current choice. However, this contradicts the notion of utility as an "ideal" index which by definition is meant to capture the value of all relevant aspects of the states.

Under suitable regularity assumptions, it follows from this definition of exogenous random preferences that the utility, $U_j(t)$ of state *j* at time *t*, viewed as a stochastic process in time (and conditional on unobservables known to the agent), becomes a so called *Extremal process* (possibly with deterministic drift). The Extremal process is analogous to the Brownian motion. Whereas the updating algorithm for the Brownian motion has the property that the realization in the current period is obtained as the sum of the realization at the previous period plus an independent increment, the realization of the Extremal process in the current period equals the maximum of the previous realization (possibly depreciated) and an independent "increment". Current utilities may depend on the utilities in the previous period due to habit persistence. The corresponding conditional choice process (given the unobservables known to the agent) turns out to be a Markov chain with a particular functional form of the transition probabilities (hazard functions) in terms of parameters of the underlying utility function.

Evidently, the assumption of the reference case is often too restrictive. We therefore discuss how the framework may be extended to allow for structural state dependence. In this case the Markov property will no longer hold. As mentioned above, this extended modeling framework allows us to separate structural state dependence effects from habit persistence provided we know the distribution of unobservables than are known to the agent.

Dagsvik (2002) has provided an alternative axiomatization of behavioral duration models. Specifically, he proposed an intertemporal version of IIA and demonstrated that it implies extremal utility processes. In contrast to Dagsvik (2002), this paper allows the utilities to be interdependent across alternatives, with multivariate extreme value distribution. Also the regularity conditions assumed by Dagsvik (2002) are considerably stronger that the ones postulated in this paper. It follows that the results of Dagsvik (2002) are special cases of the results of this paper.

Whereas the results referred to above are obtained under the assumptions of choice sets being constant over time, we also consider the case where choice sets may be non-decreasing over time. A general treatment of time varying choice sets is, however, rather complicated, and the analysis of this case is left for another occasion. However, in practice it may not be evident how interesting it is in applications to allow for time varying choice sets since one may always approximate the case where some alternatives disappear by weighting down the utilities of the respective alternatives so that they become very unattractive and consequently will almost never be chosen although they formally belong to the choice set.

In many applications it is of interest to allow for alternative-specific fixed or random effects. Recall that the model structure is derived conditional on unobservables that may sometimes be represented by random- or fixed effects, and it is therefore of particular interest to allow for such effects in our framework. We discuss identification in this case and we indicate how one can derive conditional maximum likelihood estimation procedures in these cases, similarly to Chamberlain (1980).

Another interesting challenge is the potential extension of the framework developed in this paper to models for discrete dynamic programming, similar to Rust (1987). A full treatment of the case with uncertainty and stochastic dynamic programming is, however, beyond the scope of the present paper and will be discussed elsewhere.

The paper is organized as follows: In section 2 we propose a characterization of the utility function as a random function of time (utility process) and we derive their implications for the stochastic structure of the utility process. In section 3 we derive implications for the corresponding probabilistic choice model for the case with non-decreasing choice sets. In section 5 we discuss how one can allow for structural state dependence and in section 6 we deal with identification. In section 7 we discuss how to apply the modeling framework to analyze labor supply and sectoral mobility in a static and a life cycle setting, respectively.

2. Characterization of preferences

We shall now assert assumptions with the purpose of obtaining a theoretical justification of the quantitative structure of the model. Let S be the universal set of alternative, that is, the maximal set of available alternatives, and let it be represented by the index set {1, 2, …, m}. The actual choice set B may be equal to S or be a proper subset of S. In this section we assume that the actual choice sets do not vary over time. Let $U_j(t)$ denote the agent's utility of being in state j (alternative j) at time (or age t), $j \in S$, utility function is assumed to be a random function. In the context of life cycle analysis it may be thought of as a reduced form value function representing current and future uncertain prospects. Let $J(t)$ denote the chosen alternative (state) from S at time t, $Z_j(t)$ a vector of exogenous and observable alternative-specific attributes, and $\xi_j(t)$ an exogenous random variable that captures the effect of unobservables on the utility of alternative j at time t that are perfectly known by the agent. Moreover, let $h(t) = \{J(s), s < t\}$ and $\tilde{U}(t) = \{U_j(s), s < t, j = 1, 2, ..., m\}$. That is, $h(t)$ represents the choice history prior to t and $\tilde{U}(t)$ represents all the agent's utilities prior to time t. Analogous to Heckman (1981a), a fairly general utility representation of preferences may be written as

(2.1) $$U_j(t) = \tilde{v}(Z_j(t), \xi_j(t), h(t)) + g_j(\tilde{U}(t), \eta_j(t)),$$

where $\tilde{v}(\cdot)$ and $g_j(\cdot)$ are suitable deterministic functions. Whereas $\xi_j(t)$ is known to the agent, the variable $\eta_j(t)$ is random to the agent, in the sense of Thurstone (1927). Since utility is random, the function $\{J(t), t > 0\}$ will be a random process. Thurstone's argument for allowing tastes to be random to the agent himself was that agents are viewed as having difficulties with assessing a precise and definitive value of the respective alternatives. Evidence from laboratory experiments as well as everyday observations indicate that, (i) the agents may have taste for variation, (ii) they may find it hard to assess the value of the alternatives because they may be unsure about their tastes and their perceptions may be influenced by fluctuating moods and whims, (iii) they may have limited information about the alternatives and may receive unanticipated information over time.

Let $\{U(t), t \geq 0\} \equiv \{(U_1(t), U_2(t), ..., U_m(t)), t \geq 0\}$ denote the utility function (vector of utilities) of the agent.

Assumption 1 (definition of exogenous random preferences)

Conditional on the agent's information, the value process, $\{\max_k U_k(\tau), t \geq \tau \geq s\}$, *restricted to the time interval* $[s,t]$, *is independent of the choice history prior to s,* $\{J(\tau), \tau \leq s\}$, *for any* $0 < s \leq t$.

Note that the assertion in Assumption 1 is only supposed to hold conditional on the agent's information. It asserts that the indirect utility at time *t* is independent of current and past choices. Recall first that in the standard deterministic duality theory the indirect utility is fully determined by income, prices and possible constraints. Thus, knowledge of which alternative is the most preferred does not represent additional information of relevance for the determination of the indirect utility function. Consider the stochastic setting. For simplicity, consider the special case of choice behavior at one point in time. As in the deterministic case, the indirect utility $\max_k U_k(t)$ is completely determined by preferences and choice constraints (prices, income and choice set). Thus, when preferences are exogenous, and consequently unaffected by previous choice experience, knowledge about $J(t)$ will not represent information that is relevant for assessing the value of the *level* of $\max_k U_k(t)$. If $\max_k U_k(t)$ were correlated with the choice $J(t)$ it would mean that not all information that is relevant for the agent's indirect utility is captured by the utilities, which means that the utility function would be ill defined. To further facilitate interpretation consider a large number of independent replications of a choice experiment. "Independent" means here that the respective random terms are drawn independently across experiments. In different replications the choices and the indirect utilities may be different due to different draws of the random terms. However, if the indirect utility and the choice were correlated it would mean that the c.d.f. of the indirect utility, given that state 1 (say) was chosen would be different from the corresponding c.d.f. given that state 2 (say) was chosen. This appears inconsistent with the fact that the agent's perception about the value of each alternative is fully represented by the utility function.

Assumption 2

At each given point in time t, and conditional on the agent's information, the utility process has the following properties:

(i) $$U(t) = v(t) + \varepsilon(t),$$

$v(t)$ *is a deterministic vector and* $\varepsilon(t)$ *is a random vector that is independent of* $v(t)$. *Moreover,* $v(t)$ *can attain any value in* R^m,

(ii) $\{U_1(t), t \geq 0\}$ *is independent of the processes* $\{(U_2(t), U_3(t), ..., U_m(t)), t \geq 0\}$

and

(iii) $\{U(t), t \geq 0\}$ *is continuous in probability.*

Note that the separability condition in (i) is rather weak. For example, the GEV (Generalized Extreme Value) model fulfils this condition. Dagsvik (1994) has demonstrated that within the class of random utility models the GEV model implies no (essential) restrictions on the choice probabilities. Note that the assertion that $\eta(t)$ is independent of $v(t)$ does *not* necessarily imply that the stochastic process $\{\varepsilon(t), t \geq 0\}$ is independent of the deterministic process $\{v(t), t \geq 0\}$. For example, even if $\varepsilon(t)$ and $v(t)$ are independent, $\varepsilon(t)$ may depend on $v(s)$ for some $s < t$. Using the notation of eq. (2.1) we note that $\varepsilon_j(t) = g_j(\tilde{U}(t), \eta_j(t))$. Condition (ii) does not seem to be an essential restriction due to the fact that at most utility differences can be identified (in a distributional sense), see Strauss (1979, Corollary 1, p. 40). This condition is necessary for achieving central theoretical results. However, some results obtained in this paper will continue to hold even if assertion (iii) is dropped, as we shall discuss below. Condition (iii) is a regularity property and the concept "continuity in probability" means that the probability that $|U_j(s) - U_j(t)| > \delta$, for any $\delta > 0$, tends towards zero as s tends towards t. In our context continuity in probability is a rather weak condition, and it does not imply that the sample paths of the process $\{U_j(t), t \geq 0\}$ are necessarily continuous. For example, a stochastic process with jumps may still be continuous in probability. The continuity-in-probability condition only implies that the jumps of the process cannot occur "too frequently".

Although the additive separability conditions (i) of Assumption 2 is a typical one that is often routinely invoked, it is nevertheless ad hoc from a theoretical perspective. It can however, be given a theoretical interpretation and justification, which we shall now address.

Assumption $2'$ **(***Product rule*)

Let $P_t(j,k) = P(U_j(t) > U_k(t))$. *At any point in time t, and conditional on the agent's information, the Product rule*

$$P_t(j,k)P_t(k,r)P_t(r,j) = P_t(j,r)P_t(r,k)P_t(k,j),$$

holds, for any distinct $j,k,r \in S$.

The intuition behind the product rule goes as follows: Suppose that an individual is making choice from the set $\{j,k,r\}$. Provided the choices are independent, the left hand side is the probability of the intransitive chain $j \succ k \succ r \succ j$, whereas the right hand side is the probability of the intransitive chain $j \succ r \succ k \succ j,$ where $\succ$ means "preferred to". This assertion is called the *Product rule*. The Product rule can thus be interpreted as asserting that an intransitive chain in one direction is not more probable than an intransitive chain in another direction. One can demonstrate that the Product rule holds if and only if

$$P_t(j,k) = \frac{\exp(v_j(t))}{\exp(v_j(t)) + \exp(v_k(t))},$$

for distinct $j,k \in S,$ where $\{v_j(t)\}$ are deterministic terms that are unique up to an additive constant, see Luce and Suppes (1965, p. 350). It is well known that the binary choice probabilities can be rationalized by the *additively separable* random utility model $U_j(t) = v_j(t) + \varepsilon_j(t),$ where the random error terms $\varepsilon_j(t),\ \ j \in S,$ are independent with c.d.f. $\exp(-\exp(-x)).$ This result proves that under the Product rule the utility function admits an additively separable structure. Moreover, for binary choices there is no loss of generality by letting the random error terms be independent. Thus, the Product rule has a nice theoretical interpretation (in that departure from transitivity is random) and it implies that one can represent preferences by additive random utilities.[3] Before we introduce the next assumption we need an additional definition. By *stationary environment* we mean that the observed covariates that influence the utility processes are constant over time.

Assumption 3

In a stationary environment the utility vector process converges to a stationary process as time increases.

This is an intuitively plausible assumption. The reason why we require time to be large is that there may be possible "upstarting" effects, which may gradually fade away.

The axioms above have important implication which we shall discuss below.

Theorem 1

Assumptions 1 to 3 hold if and only if the updating equation for the utility processes $\{U_j(t), t \geq 0\}$ *is given by*

(2.1) $$U_j(t) = \max\left(U_j(s) - (t-s)\theta, W_j(s,t)\right)$$

---

[3] In principle, the Product rule can also be tested non-parametrically. In the present setting such testing is however not straight forward in the case of habit persistence and unobservables.

for $0 \le s < t,$ *where* $W(s,t) \equiv (W_1(s,t), W_2(s,t), ..., W_m(s,t))$ *is independent of* $U(s) \equiv (U_1(s), U_2(s), ..., U_m(s))$ *and has standard multivariate type III extreme value c.d.f. given by*

(2.2)

$$P(W(s,t) \le x) = \exp\left(-\int_s^t H\left(\exp(w_1(\tau) - x_1), \exp(w_2(\tau) - x_2), ..., \exp(w_m(\tau) - x_m)\right) \exp(-(t-\tau)\theta) d\tau \right)$$

where θ *is a positive constant,* $w_j(t), j = 1, 2, ..., m,$ *are deterministic terms,*

(2.3)
$$H(y) = \exp(y_1) + \tilde{H}(y_2, y_3, ..., y_m),$$

$\tilde{H}$ *is a positive decreasing function on* R_+^{m-1} *with the properties;* $\tilde{H}(0,0,...,0) = 0,$ $\lim \tilde{H}(y_2, y_3, ..., y_m) = \infty,$ *when* y_j *tends towards* ∞*, for any j > 2 and for any real z,*

(2.4)
$$\tilde{H}(zy_2, zy_3, ..., zx_m) = z\tilde{H}(y_2, y_3, ..., y_m),$$

and $\tilde{H}(y_2, y_3, ..., y_m)$ *satisfies conditions that ensures that exp(-* $\tilde{H}(y_2, y_3, ..., y_m)$*) is a well defined multivariate c.d.f. Moreover,* $\{U_j(t), t \ge\}$ *is a strictly stationary Markovian process.*

Conversely, (2.1), (2.2) and (2.4) imply that Assumption 1 holds, (even if (2.3) does not hold).

The proof of Theorem 1 is given in Appendix A. Note that Theorem 1 states that the structure in (2.1), the distribution function of $W(s,t),$ as well as the independence property $W(s,t) \perp W(s',t')$, when $(s,t) \cap (s',t') = \emptyset,$ are implied by Assumptions 1 to 3. This is indeed remarkable and far from intuitive.

The term θ may be interpreted as a *preference discount* factor. This parameter plays a crucial role in the developments below because it is a key determinant of habit persistence, to be discussed further below. We note that a particular multivariate type III extreme value distribution function plays a major role here.[4] This distribution is well known within the theory of discrete choice and it implies the so-called GEV model (Generalized Extreme Value), see McFadden (1978). As regards conditions that ensure that *exp*(- $\tilde{H}(y_2, y_3, ..., y_m)$) is a multivariate c.d.f. we also refer to McFadden (1978). By standard multivariate type III extreme value distribution we mean that the corresponding marginal distributions are equal to $\exp(-\exp(-x))$.

The next result follows readily from Theorem 1.

---

[4] McFadden (1978) calls this distribution the Generalized Extreme Value distribution.

Corollary 1

Assume that time is discrete. Then under the assumption of Theorem 1 the utility functions have the structure,

(2.5) $$U_j(t) = \max\left(U_j(t-1) - \theta, w_j(t) + \eta_j(t)\right)$$

where $\eta(t) = (\eta_1(t), \eta_2(t), ..., \eta_m(t))$, *t = 1, 2, …, are independent and independent of the deterministic terms,* $\{w_j(t)\}$, *and has standard multivariate type III extreme value distributed given by*

(2.6) $$P\left(\eta(t) \le x\right) = \exp\left(-H\left(\exp(-x_1), \exp(-x_2), ..., \exp(-x_m)\right)\right),$$

where the function H satisfies (2.3) and (2.4) and necessary conditions for the right side of (2.6) to be a multivariate c.d.f., with one- dimensional univariate marginal distributions equal to [5]

(2.7) $$P\left(\eta_j(t) \le x\right) = \exp\left(-\exp\left(-x\right)\right).$$

The proof of Corollary 1 is given in the appendix.

A stochastic process with the property (2.2), and with $\theta = 0$, is called an *extremal* process in probability theory, see for example Resnick (1987). We shall call the process that satisfies (2.2) with $\theta \ge 0$, a *modified extremal process*. Thus, the modification consists in allowing for preference depreciation represented by θ. Usually, in the absence of state dependence effects, the deterministic terms $\{w_j(t)\}$ depend on time through time-dependent covariates. When θ is large there would be no dependence on the past and in this case, so that in this case $U_j(t) = w_j(t) + \eta_j(t)$. If we drop Axiom 3 then we may also allow θ to be equal to zero. Note also that in the stationary case $w_j(t)$ does not depend on *t*.

The extremal processes belong to a more general class of stochastic processes called max-stable processes. Max-stable processes have the property that they allow one to apply the maximum operation without "leaving the class". That is, the maximum of independent max-stable processes is also a max-stable process. This is analogous to the class of Gaussian processes, which is closed under aggregation. The class of modified extremal processes is also closed under the maximum operation. The modified extremal process is, possibly apart from deterministic depreciation, a pure jump stochastic process. In our context this means that the "current" utility has the role of an "anchoring" level such that unless the values of new stimuli exceed the anchoring level, utility will not be updated. The deterministic depreciation means that the "anchoring" effect at a given point in time gradually

---

[5] There seems to be some confusion in the statistical literature as regards notation. What some authors call type III extreme value distributions, others call type I. Here we have adopted the convention of Resnick (1987).

fades away as time passes. This interpretation is consistent with results from psychological research where it is typically found that individuals do respond to stimuli only if they are sufficiently strong. In fact, representations similar to (2.1) or (2.5) have a long history in psychology and measurement theory and stem from empirical evidence indicating that individuals seem not to react to stimuli unless their intensity exceeds some *sensory threshold*. This notion of sensory threshold was introduced by the philosopher Herbart (1824), see Gescheider (1997). In the present case this means that habit persistence can be interpreted as a setting where the agent does not pay attention to some "stimuli", unless the stimuli exceed some threshold determined by previous utility evaluations. Thus, with this interpretation, if $w_j(t)+\eta_j(t)\le U_j(t-1)-\theta,$ the agent will not react to the new stimulus and, accordingly, will not update her or his preferences (apart from depreciation). This interpretation is similar to Fechner's (1860/1966) notion of "just noticeable differences" (jnd). The moments of time updating occurs may not necessarily relate to actual events, but could be due to sudden glimpses of "insight" at which epochs it is realized that utility re-evaluations are needed because the value enjoyed from the respective states is not the same as it used to be. Re-evaluations may of course also happen due to unanticipated information that arrives. The special case with $\theta = 0$, corresponds to the situation in which the agent has "perfect" memory and fixed tastes in the sense that previous preference evaluations and tastes are retained perfectly fixed in the agents mind. Unless some of the systematic utility components changes, there will be no change of state in this case. The case with positive θ corresponds to the case with "imperfect" memory in the sense that previous utility evaluations are depreciated and, as a result, the current arriving stimulus will be taken into account provided $w_j(t)+\eta_j(t)>U_j(t-1)-\theta.$

Although this and similar interpretations are interesting, they are by no means crucial for the theoretical justification of our approach. It is sufficient for the rationale of our approach to rely on the intuition of Assumption 1.

Note that it follows from Theorem 1 and Corollary 1 that the utilities at any point in time are multivariate extreme value distributed. To realize this, consider for simplicity the discrete time case in which

(2.8) $$U(t)=\max_{0\le\tau\le t}\left(w(\tau)+\eta(\tau)-(t-\tau)\theta\right),$$

which by Corollary 1 implies that the c.d.f. of $U(t)$ has the form

(2.9) $$P\left(U(t)\le x\right)=\exp\left(-\sum_{\tau=0}^{t}e^{-(t-\tau)\theta}H\left(\exp(w(\tau)-x)\right)\right),$$

where $w(\tau)=(w_1(\tau),w_2(\tau),...,w_m(\tau)),\;\; x=(x_1,x_2,...,x_m)$ and

$$H(\exp(w(\tau)-x)))=H\left(\exp(w_1(\tau)-x_1),\exp(w_2(\tau)-x_2),...,\exp(w_m(\tau)-x_m)\right).$$

Clearly, the expression in (2.9) is also a multivariate extreme value distribution as is easily verified. The corresponding choice probabilities for being in a state at any given point in time follows by using the familiar formula for GEV choice probabilities, see McFadden (1978).

3. Implications for the choice probabilities in continuous time

3.1. Time-invariant choice sets

We shall now explore the implications from the theory above for the structure of the choice probabilities in the case where the choice set does not change over time, and where potential unobservable known to the agent are given. For $y \in R_+^m$ we define $H^B(y)$, obtained from $H(y)$, by setting $y_j = 0$ when $j \notin B$. (cf. McFadden, 1978).

Theorem 2

Under the assumptions of Theorem 1 it follows that $\{J(t), t > 0\}$ *is a Markov chain. The corresponding transition-and state probabilities,* $\{Q_{ij}(s,t)\}$ *and* $\{P_j(t)\}$, *can be expressed as*

(3.1)
$$Q_{ij}(s,t) = P\left(J(t) = j \middle| J(s) = i\right) = \frac{V_j(t) - V_j(s)}{\sum_{k \in B} V_k(t)},$$

for $s < j,\ j = 1,2,\ldots,m,\ i \neq j,\ j \in B,$

(3.2)
$$Q_{ii}(s,t) = 1 - \sum_{k \in B \setminus \{i\}} Q_{ik}(s,t)$$

and

(3.3)
$$P_j(t) = P\left(J(t) = j\right) = \frac{V_j(t)}{\sum_{k \in B} V_k(t)},$$

where

(3.4)
$$V_j(t) = \exp(EU_j(t) + \theta t) = \int_0^t \exp(w_j(\tau) + \theta\tau)) H_j^B\left(\exp(w(\tau))\right) d\tau + c_j,$$

and where c_j *is a positive parameter that captures the value of being in state j at time zero and* H_j^B *denotes the partial derivative with respect to component* j.

The proof of Theorem 2 follows from Dagsvik (1988), see Appendix A for details. Resnick and Roy (1990) has extended the result of Theorem 2 to the case where the joint p.d.f. of

$U(t) = (U_2(t), U_3(t), ..., U_m(t)),$ at each given point in time *t*, does not necessarily exist. This means that the function *H*(*y*) is not necessarily jointly differentiable.

Recall that the terms $\{w_j(t), t \geq 0\},$ which are treated as deterministic in this section, may depend on variables that are known to the agent by unobserved to the researcher. As mentioned above, the parameter θ is closely linked to the serial dependence of the utility processes, see Dagsvik (2002). One can show that

(3.5) $$Corr\left(\max_{k\in B} U_k(s), \max_{k\in B} U_k(t)\right) = \zeta\left(\frac{\sum_{k\in B} V_k(s)}{\sum_{k\in B} V_k(t)}\right),$$

where the function $\zeta(x), x \in [0,1]$, is given by Tiago de Oliveira (1973), and equals

(3.6) $$\zeta(x) = -\frac{6}{\pi^2}\int_0^x \frac{\log z}{1-z} dz .$$

It can be shown that $\zeta(x)$ is continuous and strictly increasing on [0,1] with $\zeta(0) = 0$ and $\zeta(1) = 1$. Alternatively, one can also show that (Resnick and Roy, 1990),

(3.7) $$Corr\left(\exp(-\max_{k\in B} U_k(s)), \exp(-\max_{k\in B} U_k(t))\right) = \frac{\sum_{k\in B} V_k(s)}{\sum_{k\in B} V_k(t)}.$$

As measures of dependence the expressions in (3.5) and (3.7) are equivalent since they differ only by a strictly increasing transformation. From (3.4) it follows that when $\sum_{k\in B} V_k(s) / \sum_{k\in B} V_k(t) \to 0,$ then Hence, the corresponding autocorrelation in (3.5) and (3.7) tend towards zero in this case, such that in the limit there is no serial correlation in tastes. It follows readily from Theorem 2 that in this case $Q_{ij}(s,t) = P_j(t)$ where

(3.8) $$P_j(t) = \frac{\exp(w_j(t)) H_j^B\left(\exp(w(t))\right)}{\sum_{k\in B} \exp(w_k(t)) H_k^B\left(\exp(w(t))\right)}.$$

The expression in (3.8) is the familiar formula for the choice probabilities of the GEV family, see McFadden (1978). In the other extreme case when $\theta \to 0,$ then $\sum_{k\in B} V_k(s) / \sum_{k\in B} V_k(t) < 1$ for $s < t,$ and finite for *t,* but as *t* tends towards infinity this ratio tends towards 1, which corresponds to perfect serial dependence in tastes.

From Theorem 2 we see that an important implication of (modified) extremal utility processes is that that the choice process $\{J(t), t > 0\}$ is a *Markov chain* in continuous time (this holds in the general case with non-stationarity). Thus, in the "reference" case with no state dependence effects Axiom 2 provides a theoretical motivation for assuming the Markov property. In addition, Assumptions 1 to 3 imply a particular structure of the transition- and state probabilities, as expressed in (3.1) to (3.4).

The next result follows immediately from Theorem 1.

Corollary 2

Suppose that S contains at least three alternatives. When the choice set contains more than two alternatives it follows from Theorem 1 that the choice probability given a transition equals

(3.9)
$$\pi_{ij}(t)=\frac{V_j'(t)}{\sum_{k\in B\setminus\{i\}}V_k'(t)}=\frac{\exp(w_j(t))H_j^B\left(\exp(w(t))\right)}{\sum_{k\in B\setminus\{i\}}\exp(w_k(t))H_k^B\left(\exp(w(t))\right)}.$$

We recognize the formula in (3.9) as a GEV choice model, that is, the choice probability that follows from maximizing a utility function of the form $w_j(t)+\eta_j(t),$ subject to the choice set $B\setminus\{i\},$ where $(\eta_1(t),\eta_2(t),...)$ are multivariate extreme value distributed.

From Corollary 2 it follows moreover that for

Corollary 3

Suppose that S contains at least three alternatives. Under the assumptions of Theorem 1 the transition probabilities $\{Q_{ij}(s,t)\}$ *of the Markov chain* $\left\{J(t),t>0\right\}$ *have the property that* $Q_{ij}(s,t)=Q_{rj}(s,t),$ *when* $i\neq r,i\neq j,r\neq j.$

The result in Corollary 3 is rather intriguing: It asserts that under the absence of state dependence, the transitions to a new state are stochastically independent of the state of origin. Thus, this property is in fact a characterization of the reference case of no state dependence.

Next we shall consider the corresponding transition intensities. Recall that the two state case the hazard rates of a continuous time Markov chain are defined (usually) as
$$q_{ij}(t)=\lim_{s\to t}\frac{Q_{ij}(s,t)}{t-s},$$
for $i\neq j.$

Corollary 4

Let $T_i(s)$ *be the duration of stay in state i, given that state i was entered at time s. Under the conditions of Theorem 1 the transition intensities of the Markov chain* $\left\{J(t),t>0\right\}$ *are given by*

(3.10) $$q_{ij}(t)=\frac{V_j'(t)}{\sum_{k\in B}V_k(t)}=\frac{\exp(w_j(t))H_j^B\left(\exp(w(t))\right)}{\sum_{k\in B}V_k(t)},$$

for $i\neq j,\ j\in B,$ *and*

(3.11) $$q_{ii}(t)=\sum_{k\in B\setminus\{i\}}q_{ik}(t).$$

Furthermore,

(3.12) $$P\left(T_i(s)>y\right)=\exp\left(-\int_s^{s+y}q_{ii}(\tau)d\tau\right).$$

The intensity $q_{ii}(t)$ is the *hazard function* (hazard rate) at time *t.* The formula in (3.12) for the duration c.d.f. can be simplified a bit. For notational simplicity, write

$$V(t)=\sum_{k\in B}V_k(t).$$

With this notation we can express the hazard function as

(3.13) $$q_{ii}(t)=\frac{V'(t)-V_i'(t)}{V(t)}.$$

When the expression in (3.13) is inserted into (3.12) we obtain that

$$P\left(T_i(s)>y\right)=\frac{V(s)}{V(s+y)}\exp\left(\int_s^{s+y}\frac{V_i'(\tau)}{V(\tau)}d\tau\right).$$

Example 3.1

In this example we consider a choice setting with 3 alternatives, i.e., S = {1,2,3}. The error terms of the utilities of alternatives 2 and 3 are allowed to be correlated. To this end we assume that the joint c.d.f. of the error terms of the increment of utility function, $(\eta_1(t),\eta_2(t),\eta_3(t)),$ is given by

$$P(\eta_1\le x_1,\eta_2\le x_2,\eta_3\le x_3)=\exp(-e^{-x_1}-(e^{-x_2/\rho}+e^{-x_3/\rho})^{\rho}),$$

where $\rho\in(0,1].$ This specification implies that

$$corr(\varepsilon_2,\varepsilon_3)=1-\rho^2,$$

and

$$H(x)=e^{x_1}+(e^{x_2/\rho}+e^{x_3/\rho})^{\rho}.$$

Hence, it follows that

$$V_1(t)=\int_0^t\exp(w_1(\tau)+\tau\theta)+c_1,\qquad V_1'(t)=\exp(w_1(t)+t\theta),$$

$$V_j(t)=\int_0^t e^{\theta\tau}\left(\exp(w_2(\tau)/\rho+\exp(w_3(\tau)/\rho\right)^{\rho-1}\exp(w_j(\tau)/\rho)d\tau+c_j,$$

and

$$V_j'(t)=e^{\theta t}\left(\exp(w_2(t)/\rho+\exp(w_3(t)/\rho\right)^{\rho-1}\exp(w_j(t)/\rho),$$

for j = 2, 3. The corresponding transition intensities now follow by inserting the expressions above in (3.10). Similarly, the distribution of the holding times in the respective states follows from (3.11).

Corollary 5

In the special case where the utility processes are independent Theorem 2 implies that

$$V_j(t)=\int_0^t \exp\left(w_j(\tau)+\tau\theta\right)d\tau+c_j.$$

Furthermore, if B contains at least three alternatives, the choice probabilities, given transition out of the state occupied reduces to

(3.14) $$\pi_{ij}(t)\equiv P\left(J(t)=j\mid J(t-)=i,J(t)\neq J(t-)\right)=\frac{\exp(w_j(t))}{\sum_{k\in B\setminus\{i\}}\exp(w_k(t))}.$$

We note that the conditional transition probability in Corollary 5, satisfies the IIA property, see Luce (1959). From Corollary 5 the following result follows immediately:

Corollary 6

Under the conditions of Corollary 5 the transition intensities of the Markov chain $\{J(t),t>0\}$ *are given by*

(3.15) $$q_{ij}(t)=\frac{\exp\left(w_j(t)\right)}{\sum_{k\in B}\left(\int_0^t\exp\left(w_k(\tau)-(t-\tau)\theta\right)d\tau+e^{-\theta t}c_k\right)},$$

for $i\neq j$, *and for* $i=j$,

(3.16) $$q_{ii}(t)=\sum_{k\in B\setminus\{i\}}q_{ik}(t).$$

The corresponding transition probabilities and the unconditional choice probabilities are given by

(3.17) $$Q_{ij}(s,t)=\frac{\int_s^t \exp(w_j(\tau)-(t-\tau)\theta)d\tau}{\sum_{k\in B}\left(\int_0^t \exp(w_k(\tau)-(t-\tau)\theta)d\tau+e^{-t\theta}c_k\right)},$$

for $i\neq j$, *and*

(3.18) $$P_j(t)=\frac{\int_0^t \exp(w_j(\tau)-(t-\tau)\theta)d\tau+e^{-\theta t}c_j}{\sum_{k\in B}\left(\int_0^t \exp(w_k(\tau)-(t-\tau)\theta)d\tau+e^{-t\theta}c_k\right)}.$$

In Appendix B we have outlined how the transition probabilities and intensities look like in the case where the deterministic functions $\{w_j(t)\}$ are step functions that only changes at discrete time periods. In the case with independent utility processes across alternatives and with time independent systematic utility components $\{w_j(t)\}$, and $c_j=0$, for all *j*, the result of Corollary 6 reduces to

(3.19) $$P_j(t)=P_j=\frac{\exp(w_j)}{\sum_{k=1}^{m}\exp(w_k)}$$

and

(3.20) $$Q_{ij}(s,t)=P_j\cdot\frac{1-\exp(-(t-s)\theta)}{1-\exp(-t\theta)}$$

for $s<t$, $i\neq j$, and

(3.21) $$Q_{ii}(s,t)=\frac{\exp(-(t-s)\theta)-\exp(-t\theta)}{1-\exp(-t\theta)}+P_i\cdot\frac{1-\exp(-(t-s)\theta)}{1-\exp(-t\theta)}.$$

The autocorrelation function in this case reduces to

(3.22) $$Corr(U_j(s),U_j(t))=\zeta\left(1-\frac{1-\exp(-(t-s)\theta)}{1-\exp(-t\theta)}\right).$$

The corresponding transition intensities are given by

(3.23) $$q_i(t)=\frac{\theta P_j}{1-\exp(-t\theta)}.$$

From (3.19) to (3.23) we see that the transition probabilities and intensities become stationary when $t\theta$ is large. However, when $t\theta$ is small this is not so. In fact, the transition probabilities and intensities increase by the factor $1/(1-\exp(-t\theta))$ when $t\theta$ is small. The interpretation is that when the agent is very "young" the choice history is very short and therefore the effect of habit persistence is weak. As the agent grows older this "upstarting" effect disappears gradually, and becomes negligible when $\exp(-t\theta)$ is close to zero. This is also seen from the autocorrelation function in (3.22). When $t\theta$ is large, the autocorrelation function is, apart from a strictly increasing transformation, equal to $\exp(-t\theta)$. However, when $t\theta$ is "small" the autocorrelation is influenced by the term $1/(1-\exp(-t\theta))$ in such a way that it becomes weaker when t decreases. Consider the special case when θ is close to zero. Then (3.20) and (3.21) are approximately equal to

$$Q_{ij}(s,t) = \frac{t-s}{t} \cdot P_j,$$

for $i \neq j$, and

$$Q_{ii}(s,t) = \frac{s}{t} + \frac{t-s}{t} \cdot P_i.$$

In this case the autocorrelation function becomes,

$$Corr\left(\max_k U_k(s), \max_k U_k(t)\right) = \zeta\left(\frac{s}{t}\right) = \zeta\left(1 - \frac{(t-s)}{t}\right).$$

Thus, when time is large there are no transitions in this case. However, when t is "small" the autocorrelation will be less than 1 which means that in the beginning the effect of habit persistence is reduced because the choice history is short.

In the special case where $w_j(t)$ does not depend on *t* and *t* is large the transition probabilities and intensities reduce to

(3.24) $$Q_{ij}(s,t) = P_j\left(1 - \exp\left(-(t-s)\theta\right)\right)$$

for $s < t$, $i \neq j$, and

(3.25) $$Q_{ii}(s,t) = \exp\left(-(t-s)\theta\right) + P_i\left(1 - \exp\left(-(t-s)\theta\right)\right).$$

The corresponding transition intensities, for $i \neq j$, are given by

(3.26) $$q_{ij}(t) = \theta P_j.$$

The depreciation effect represented by the parameter θ can in fact be given an interesting interpretation, which we shall now explain. In fact, one can interpret the depreciation mechanism as a stochastic device where the habit persistence effect is represented by means of a particular Poisson process. The intuition is that, at independent random points Z_k in time, $k = 1, 2, \ldots$, (random to the observer), the agent forgets, or stops, caring about previous evaluations and only takes into account and new stimuli. That is, if for some *k*, $Z_{k-1} < t < Z_k$, the agent will, at time *t*, only take into account previous preference evaluations within the interval $\left(Z_{k-1}, t\right]$. The intensity of this process is θ, which means that the probability that for some *k*, $Z_k \in (t, t+dt)$ equals θdt. If the agent *does* care, he will keep the previous value *without* depreciation. Consider now (3.24). Since no explanatory variable change over time, the agent will never change to a new alternative unless he stops taking the past into account. Specifically, to change state from time *s* to time *t* the agent must stop taking into account previous evaluations of the alternatives at some time in the interval (s,t). Since the interval between two events in a Poisson process is exponentially distributed with θ as parameter, the probability that this will happen is $1-\exp(-(t-s)\theta)$. The conditional probability that the agent shall choose alternative *j*, given that he stops caring about the past, equals P_j, since the corresponding utilities that govern this choice are $U_j = w_j + \eta_j(t)$, where $\eta_j(t)$, $j = 1, 2, ..., m$, are independent extreme value distributed as in (2.7). Hence, by multiplying P_j by the probability that the agent will begin to neglect the past some time within (s,t), (3.24) is obtained. Similarly, if the agent occupies state *i* at time *s* he will continue to be in state *i* at time *t* if the agent cares about past preference evaluations (which has probability equal to $\exp(-\theta(t-s))$, or if the agent does not care about past evaluations (with probability $1-\exp(-\theta(t-s))$) but chooses state *i* with probability P_i. Hence, (3.25) follows. Eq. (3.26) asserts that the intensity of a transition from *i* to *j* can happen when the agent stops taking the past into account in $\left(t, t+dt\right)$, which happens with probability θdt. Given that he forgets about the past he will go to state *j* with probability P_j. Hence, we obtain that the probability of going from state *i* to *j* in $\left(t, t+dt\right)$ equals $\theta dt P_j$.

Let us compare the structure of the preferences and their implications obtained above with a formulation based on a multiperiod Probit framework. For simplicity we shall only consider the binary case. To this end assume that

(3.27) $$U_j(t) = \rho U_j\left(t-1\right) + w_j(t) + \eta_j(t)$$

where $\eta_j(t)$, $j = 1, 2, t = 1, 2, ...$, are i.i. normally distributed random variables with variance equal to 1/2. The structure in (3.27) is analogous to the extremal process given in (2.5). In particular, both representations have the Markov property. It follows easily that

(3.28) $$P\left(U_2(t) > U_1(t)\right) = \Phi\left(\sqrt{\frac{1-\rho^2}{1-\rho^{2(t+1)}}}\left(\sum_{k\le t} \rho^k \left(w_2\left(t-k\right) - w_1\left(t-k\right)\right)\right)\right).$$

Not surprisingly, the structure in (3.28) is similar to the one obtained in Corollary 6 for the binary case. The corresponding transition probabilities cannot, however, be expressed on closed form. The multiperiod probit setting with utility as in (3.27) has been discussed by Heckman (1981a, section 3.10).

Although the specification in (3.27) and the choice probability in (3.28) are analogous to (2.5) and (3.18) the implications for the choice process over time are qualitatively different. Whereas the extremal process representation of the utility processes that follows from Assumptions 1 to 3 yields a *Markovian* choice process, the first order autoregressive Gaussian process representation in (3.27) does *not* imply a Markovian choice process. However, the most important weakness of the multiperiod Probit model is the lack of choice theoretic rationale for the model structure (in the reference case).

4. Non-decreasing choice sets

Here, we shall consider the situation with time varying choice sets. In general, this case is more complicated and we shall limit our treatment to the case in which the utility processes are independent across alternatives and where the choice sets, $\{B(t), t > 0\}$ are non-decreasing, which means that $B(s) \subseteq B(t)$, whenever $s < t$. We have the following result.

Theorem 3

Assume that $B(s) \subseteq B(t)$, whenever $s < t$ and that the alternative-specific utility processes are independent across alternatives. Then the choice process $\{J(t), t > 0\}$ is a Markov chain with transition- and state probabilities given by

$$Q_{ij}\left(s,t\right) = P\left(J(t) = j \middle| J(s) = i\right) = \frac{V_j(t) - \delta_j(B(s))V_j(s)}{\sum_{k \in B(t)} V_k(t)},$$

for $s < j,\ j = 1,2,...,m,\ i \neq j, i \in B(s),\ j \in B(t),$

$$Q_{ii}\left(s,t\right) = 1 - \sum_{k \in B(t)\setminus\{i\}} Q_{ik}\left(s,t\right).$$

and

$$P_j(t) = P\left(J(t) = j\right) = \frac{V_j(t)}{\sum_{k \in B(t)} V_k(t)},$$

where $\delta_k(B) = 1,$ *if k belongs to the set B and zero otherwise, and*

$$V_j(t) = \int_0^t \exp(w_j(\tau) + \theta\tau) d\tau + c_j.$$

Theorem 3 is proved in Appendix A. From Theorem 3 the next result is immediate;

Corollary 7

If $j \in B(t) \setminus B(s)$, *then*

$$P\big(J(t) = j \,\big|\, J(s)\big) = P(J(t) = j). \tag{4.1}$$

The result of Corollary 7 asserts that the choice of "new" alternatives, that is, alternatives that are available at the current time epoch but were not available previously, is independent of previous choices.

Dagsvik (2002) postulated (4.1) as an axiom that should hold in the absence of structural state dependence, and he called it an Intertemporal version of the Independence from Irrelevant Alternatives, property (IIIA). He also proved that when the utilities $\{U_j(t), t \geq 0\}$, j =1, 2,…, are independent max-stable processes then (4.1) implies that the utilities are Extremal processes. Consequently, Corollary 7 follows in this case.

Unfortunately, the case with general time varying choice sets is rather complicated, and will be discussed elsewhere.

5. Structural state dependence

In the analysis of panel data it is often noted that individuals who have experienced an event in the past are more likely to experience the event in the future than are individuals who have not experienced the event. Recall that there are two explanations of this: One explanation is that as a consequence of experiencing the event preferences and choice restrictions are altered. If so, otherwise identical individuals who did experience the event would behave differently in the future than individuals who did not experience the event. In this case choice experience has a genuine behavioral effect that give rise to so-called true state dependence. Another explanation is that preference and choice constraints differ across individuals in an unmeasured way. This implies that previous and future choices will be correlated, since choices represent information about temporally persistent unobservables that affect choices. Thus, this case yields so-called spurious state dependence, (Heckman, 1981a,b). As Heckman has emphasized in several papers, one cannot without a priori theoretical restrictions separate spurious from true state dependence. In other words, this identification problem cannot be settled by statistical methods alone (see for example Heckman, 1991).

In this case the structural part of the utility function becomes endogenous and it is far from evident how the corresponding choice probabilities should be specified in this case.

Suppose initially that no state dependence were present such that preferences were exogenous and Assumption 1 to 3 hold. Then the utility representation in (2.5) holds. Now suppose that preferences may be affected by choice behavior. This means that either the systematic terms $\{w_j(t)\}$ or the distribution of the random error terms $\{\eta_j(t)\},$ or both, are influence by choice behavior. Let $h(t) = \{J(s), s < t\}$ denote the choice history prior to time *t* and let $w_j(t, h(t))$ and $\theta(h(t))$ denote the structural term of utility of alternative *j* and habit persistence parameter, respectively, modified by the choice history.

Assumption 4

Under structural state dependence the utility function is updated according to

$$U_j(t, h(t)) = \max(U_j(t-1, h(t-1)) - \theta(h(t)), w_j(t, h(t)) + \eta_j(t))$$

where the vectors $\eta(t), t = 1, 2, ...,$ *are independent of the choice history, and with distributional properties as in Theorem 1.*

Assumption 4 asserts that when the effect of previous choice experience is properly accounted for, in the sense that possible behavioral effects on the parameters of the utility function have been incorporated, then there is no interdependence between the current and future indirect utility and previous choices. The motivation for this is analogous to the justification of Assumption 1. Indeed, if this were not true it would signify that not all the relevant information about the value of the alternatives has been accounted for in the utility index. In other words, in this case the indirect utility would be ill defined since it fails to capture all information relevant for the agent's choice behavior.

The next result is an implication of Assumption 4.

Theorem 4

Assume that Assumptions 1 to 4 hold. Then the discrete time version of the transition probabilities hold with $\{w_j(s), s < t\}$ *replaced by the corresponding experience dependent terms,* $\{w_j(s, h(s)), s < t\}.$

The proof of Theorem 4 is given in Appendix A.

The result of Theorem 4 means that one can introduce previous choice variables into the transition probabilities or transition intensities as if they were exogenous. However, the nice formula for the probability of being in a particular state will no longer hold.

6. Identification and conditional estimation

In this section we shall discuss identification. In particular, we shall consider identification in three separate cases, namely the case with fixed effects, random effects and the case with neither random nor fixed effects. For simplicity, we assume throughout this section that the choice sets are constant over time and that the utility processes are independent over alternatives. In general, it may not be possible to represent the unobservables (known to the agent) by random or fixed effects because they may not be constant over time. The more general case with non-constant unobservables will not be discussed here.

6.1. Identification in the absence of unobservables that are perfectly known to the agent.

In this subsection we shall discuss identification in the basic case where there are no unobservables that are perfectly known by the agent. In particular, we shall show that given the transition intensities the respective partial derivatives of the *H* function can be recovered. For simplicity, the next result is given only for the case with exogenous preferences (no state dependence), although it is possible to extend the result to accommodate state dependence.

Corollary 8

Let A be any proper subset of the choice set B. Under the assumptions of Theorem 2 it follows, for $i \in B \setminus A, j \in A,$ *that*

$$\log\left(\int_d^t H^B\left(w(\tau)\right)e^{\theta\tau}d\tau+\sum_{k\in B}c_k\right)+\theta t=\int_d^t\left(\sum_{k\in A}q_{ik}(x)+\sum_{k\in B\setminus A}q_{jk}(x)\right)dx+\kappa, \tag{6.1}$$

and

$$\log H_j^B\left(\exp(w(t))\right)+w_j(t)+\theta t=\log q_{ij}(t)+\int_d^t\left(\sum_{k\in A}q_{ik}(x)+\sum_{k\in B\setminus A}q_{jk}(x)\right)dx+\kappa, \tag{6.2}$$

where $d \ge 0$ *and* κ *are arbitrary constants.*

The proof of Corollary 8 is given in appendix A.

Corollary 8 shows that $H_j^B(\exp(w(t)))\exp(w_j(t)+t\theta)$ is non-parametrically identified up to an arbitrary multiplicative constant, κ. This constant is irrelevant because it cancels in the choice model. In particular (6.2) implies that

(6.3) $$w_1(t) + \theta t = \log q_{i1}(t) + \int_d^t \left(\sum_{k \in A} q_{ik}(x) + \sum_{k \in B \setminus A} q_{1k}(x) \right) dx + \kappa.$$

From the transition intensities we have that

(6.4) $$\log H_j^B\left(\exp(w(t))\right) + w_j(t) - w_1(t) = \log q_{ij}(t) - \log q_{i1}(t)$$

for $i \neq j, 1.$ The relation in (6.4) are analogous to the corresponding identifying relations between the utility structure and the choice probabilities in the GEV model. Under suitable restrictions the structural terms $\{w_j(t)\}$ can be identified. Consequently, (6.3) implies that θ is identified.

Consider the particular case with only two alternatives and with utility processes that are independent across alternatives. In this case (6.3) reduces to

(6.5) $$w_1(t) + \theta t = \log q_{21}(t) + \int_d^t \left(q_{12}(x) + q_{21}(x) \right) dx + \kappa$$

and

(6.6) $$w_2(t) + \theta t = \log q_{12}(t) + \int_d^t \left(q_{12}(x) + q_{21}(x) \right) dx + \kappa .$$

In contrast to the case with more than two alternatives, there are only two transition intensities in the binary case. Thus for any pair of intensity functions $q_{12}(t), q_{21}(t)$ we can find a pair $w_1(t)$ and $w_2(t)$ determined by (6.5) and (6.6). In other words, this implies that *any* Markov chain with two states has an independent Extremal process utility representation. We summarize this property in the corollary below.

Corollary 9

A Markov chain in continuous time with two states always admits an independent Extremal utility process representation.

6.2. Identification with additive fixed effects and state dependence

We assume in this section that the random error terms of the utility functions are independent across alternatives and that the respective structural terms contain fixed- or random effects. Consider first the case with fixed effects and assume that

(6.7) $$w_j(t, h(t)) = \mu_j + \tilde{w}_j(t, h(t)),$$

where μ_j and $\tilde{w}_j(t, h(t)),$ j = 1,2,…, are, respectively, fixed effects and deterministic functions of individual characteristics and alternative-specific attributes that are assumed known apart from a set of parameters to be estimated. In general, the structural terms $\{w_j(t,h(t))\}$ may depend on choice experience through $h(t)$. We normalize by letting $\mu_1 = 0.$ From Corollary 6 it now follows immediately that

(6.8) $$\log\left(\frac{q_{ij}(t,h(t))q_{i1}(s,h(s))}{q_{i1}(t,h(t))q_{ij}(s,h(s))}\right) = \tilde{w}_j(t,h(t)) - \tilde{w}_1(t,h(t)) + \tilde{w}_1(s,h(s)) - \tilde{w}_j(s,h(s)),$$

for $i \neq j,1,$ and $j \neq 1.$ Eq. (6.8) demonstrates that when covariates change over time one is able to identify

$$\tilde{w}_1(s,h(s)) - \tilde{w}_j(s,h(s)) + \tilde{w}_j(t,h(t)) - \tilde{w}_1(t,h(t)).$$

If, for example, $\tilde{w}_j(t,h(t)) = Z_j(t,h(t))\beta$ where $\{Z_j(t,h(t))\}$ are K-dimensional attribute vectors that also may be individual-specific, and possibly modified by choice experience. This implies that

(6.9) $$\tilde{w}_1(s) - \tilde{w}_j(s) + \tilde{w}_j(t) - \tilde{w}_1(t) = \sum_{k=1}^{K}(Z_{1k}(s) - Z_{jk}(s) + Z_{jk}(t) - Z_{1k}(t))\beta_k.$$

Thus, with standard conditions on the matrix $\{Z_{1k}(s) - Z_{jk}(s) + Z_{jk}(t) - Z_{1k}(t)\},$ it follows that the parameter vector β is identified.

6.3. Conditional likelihood estimation with fixed and random effects

Assume now that there is no state dependence. Thus (6.7) reduces to $w_j(t) = \mu_j + \tilde{w}_j(t),$ where μ_j is a fixed or random effect. Consider for simplicity the case with discrete time and let $q_{ij}(t)$ be the one step transition probabilities from time *t-1* to *t*. Chamberlain (1980) and Lee (2002) have demonstrated that one can form a particular conditional likelihood function in multinomial logit models where the fixed effects are eliminated. The duration model we have analyzed in this paper has a similar property. Specifically, it follows that

(6.10) $$\frac{q_{1j}(t-1)q_{j1}(t)}{q_{1j}(t-1)q_{j1}(t) + q_{j1}(t-1)q_{1j}(t)} = \frac{1}{1+\exp(\tilde{w}_1(t-1) - \tilde{w}_j(t-1) + \tilde{w}_j(t) - \tilde{w}_1(t))}.$$

The expression in (6.10) is the conditional likelihood of transitions from state 1 at time *t*-2 to state j at time *t*-1 and further back to state 1 at time *t*, given that transitions occur either from state *i* to state *j*, or vice versa. Thus, similarly to Chamberlain's conditional likelihood method and the results above imply the potential of developing a similar conditional likelihood method for estimating the structure of $\{\tilde{w}_j(t)\}$ in a similar way as Honoré and Kyriazidou (2000). Unfortunately, in the presence of state dependence one cannot apply this method.

Consider next the analogous case with random effects. Let

(6.11) $$q_{ij}(t,h(t)\,|\,\mu)=\frac{\exp(\mu_j+\tilde{w}_j(t,h(t)))}{V(t,h(t)\,|\,\mu)e^{-\theta t}},$$

where $\mu=(\mu_1,\mu_2,...\}$ and

$$V(t,h(t)\,|\,\mu)=\sum_k\left[\int_0^t \exp(\mu_k+\tilde{w}_k(\tau,h(\tau))+\tau\theta)d\tau+c_k\right].$$

From (6.11) it follows that the unconditional transition probability of going from state i at time t-2 to state j at time t-1 and back to state i at time t, $i\neq j$, is equal to

$$E\left(q_{ij}(t-1\,|\,\mu)q_{ji}(t\,|\,\mu)\right)=\exp(\tilde{w}_j(t-1)+\tilde{w}_i(t))E\left(\frac{\exp(\mu_i+\mu_j)}{V(t-1\,|\,\mu)V(t\,|\,\mu)e^{-(2t-1)\theta}}\right),$$

which implies that

(6.12) $$\frac{E\left(q_{1j}(t-1\,|\,\mu)q_{j1}(t\,|\,\mu)\right)}{E\left(q_{1j}(t-1\,|\,\mu)q_{j1}(t\,|\,\mu)\right)+E\left(q_{j1}(t-1\,|\,\mu)q_{j1}(t\,|\,\mu)\right)}=\frac{1}{1+\exp(\tilde{w}_1(t-1)-\tilde{w}_j(t-1)+\tilde{w}_j(t)-\tilde{w}_1(t))}.$$

Similarly to the case with fixed effects, the expression in (6.12) is the conditional likelihood of transitions between state 1 and state j, given that transitions occur between state 1 to state j, or vice versa. Thus, also in the case with random effect it is possible to apply a conditional likelihood method for the estimation of $\{\tilde{w}_j(t)\}$ that does not depend on the postulated distribution of the random effects. This is clearly a great advantage since it is hard to justify a priori the class of distributions of the random effects.

7. Further Examples

In this section we shall briefly discuss how the stochastic framework developed in this paper might be applied in concrete cases.

7.1. Labor Supply and Sectoral mobility in a static setting

In this example the issue is to analyze panel data on sectoral mobility when one allows for habit persistence. The agent is assumed to have the choice between different sectors in the labor market. Each sector has fixed (sector-specific) hours of work and wage. Let z_{tj} denote the wage at time period t (discrete time) of sector j. There are no transaction costs and the agent chooses to be in the state that maximizes utility. Assume that the utility of working in sector j at t, with disposable income C and workload h, has the structure $u_t(C,h)+\gamma_j+\varepsilon_j(t)$, where $\{\varepsilon_j(t)\}$ are random error terms, $u_t(C,h)$ is a deterministic term that captures the representative utility of disposable income C and hours of work h, and γ_j is a deterministic term that captures the representative utility of the tasks to be performed in occupation *j*, ceteris paribus. For sector *j* hours of work is fixed and equal to h_j. Given sector *j*, the budget constraint is given by $C = f(z_{tj}h_j, I),$ where *I* is non-labor income and *f* is the function that transforms gross income $z_{tj}h_j + I$ to disposable income. Let

(7.1) $$w_j(t) = u_t(f(z_{tj}h_j, I_j), h_j) + \gamma_j.$$

Thus, in the absence of taste persistence we interpret $w_j(t) + \eta_j(t)$ is the conditional indirect utility function at time *t*, given alternative *j*. The term $w_j(t)$ is the period-specific representative utility of occupation *j* at time *t*. The term γ_j could be specified as a function of non-pecuniary and sector-specific attributes. It could also be treated as fixed or random effect. If we allow for habit persistence and invoke Assumptions 1 to 3 it follows from Corollary 1 that the utility function is given by

(7.2) $$U_j(t) = \max(U_j(t-1) - \theta, w_j(t) + \eta_j(t)),$$

where the vector $\eta(t) = (\eta_1(t), \eta_2(t), ...)$ is serially independent and multivariate extreme value distributed. The probabilities for going from one occupation to another one, and the probability of working in a particular occupation are now given by the formulas in Theorem 2 or Corollary 6 in the case with independent utilities. If γ_j is a fixed or random effect one could in fact estimate the model by the method suggested in sections 6.1 and 6.2.

7.2. Labor supply and sectoral mobility in a life cycle setting

In this section we shall discuss the modeling of labor supply and sectoral choice and mobility in a life cycle context, where we in addition allow for habit persistence. The approach is analogous to Heckman and McCurdy (1980). As in the previous example we assume that time is discrete. However, the corresponding continuous time version can be treated similarly. In this section we assume that the agent is bounded rational in the sense that she or he is uncertain about the future tastes, interests and wages and is only able to make predictions about the level of these variables but may revise plans at each epoch due to unanticipated information. Let C_t denote consumption in period *t* and z_{tj} and h_{tj} the wage and hours of work in sector j in period *t*. Furthermore, let A_t be the wealth at the end of period $t-1$, and *r* the interest rate. Length of life is *T*. Conditional on occupation *j* in period *t* the agent faces the budget constraint

(7.3) $$A_{t+1} = \sum_{j \in B_t} h_{tj} z_{tj} + (1+r) A_t - C_t.$$

Let $\boldsymbol{h}_t = (h_{t1}, h_{t2}, ..., h_{tm})$. Assume that the life time utility is additive separable with period-specific utility given by

(7.4) $$u_t(C_t, \boldsymbol{h}_t) = u_{t1}(C_t)\xi_{t1} + u_{t2}\left(\sum_k h_{tk}/\gamma_{tk}\right)\xi_{t2}$$

where $\{\xi_{t1}, \xi_{t2}, \gamma_{tk}\}$ are positive random taste shifters and u_{t1} and u_{t2} are respectively strictly increasing and decreasing, concave differentiable functions. The taste shifters are assumed to be realizations of random variables. At time *t*, current and future taste-shifters $\{\xi_{\tau 1}, \xi_{\tau 2}, \gamma_{\tau k}, T \geq \tau \geq t\}$ are known by the agent, but at the next point in time *t*+1, the agent draws new current and future taste-shifters $\{\xi_{\tau 1}, \xi_{\tau 2}, \gamma_{\tau k}, T \geq \tau \geq t+1\}$. How these taste-shifters are distributed and serially correlated will be discussed in a moment. Thus, similarly to the previous setup in this paper, the agent is assumed to be uncertain about his preference evaluations and may shift her or his tastes over time in a manner that is not fully foreseeable to her or him. As regards future interest rates, wages and prices the agent behave in each period as if they were known with perfect certainty but may adjust these at the next moment in time. The structure in (7.4) implies that at each moment in time, the occupations are perfect substitutes in the sense that by suitably adjusting the respective workloads the jobs will be equally attractive, provided the wages are equal across occupations. The agent's optimization problem in period *t* is to maximize

$$u_t\left(C_t, \boldsymbol{h}_t\right) + \beta \tilde{V}_{t+1}\left(A_{t+1}\right)$$

subject to (7.3), given A_t, where $\tilde{V}_t\left(A_t\right)$ is the value function given the agent's information at period *t*. To facilitate optimization it is convenient to reformulate the problem as follows: Let $x_{tj} = h_{tj}/\gamma_{tj}$ and $\tilde{z}_{tj} = \gamma_{tj} z_{tj}$. With this notation the optimization problem consists in maximizing

(7.5) $$u_{t1}(C_t)\xi_{t1} + u_{t2}\left(\sum_{j \in B_t} x_{tj}\right)\xi_{t2} + \beta \tilde{V}_{t+1}\left(A_{t+1}\right)$$

subject to

$$A_{t+1} = \sum_{j \in B_t} x_{tj}\tilde{z}_{tj} + \left(1+r\right)A_t - C_t .$$

Since the modified workloads, represented by $\{x_{tj}\}$, enter symmetrically in (7.5), it follows that the agent will prefer to work at most in one sector, namely in the sector with the highest value of the modified wage rates $\{\tilde{z}_{tj}\}$. In other words, the index *J* of the chosen sector is determined by

$$J_t = j \text{ if } \tilde{z}_{tj} = \max_{k \in B_t} \tilde{z}_{tk} .$$

Furthermore, consumption, hours of work and wealth are determined such that

$$u'_{t1}(C_t)\xi_{t1} = \lambda_t$$

and

$$-u'_{t2}(x_{tJ})\xi_{t2} \geq \lambda_t \tilde{z}_{tJ},$$

where λ_t is the marginal utility of wealth as of period (year) *t*. In particular, the agent will not work in period *t* if

$$-\lambda_t^{-1} u'_{t2}(0)\xi_{t2} > \tilde{z}_{tJ} = \max_{k \in B_t} \tilde{z}_{tk}.$$

The marginal utility of wealth is determined by the Euler condition

(7.6) $$\lambda_t = \beta(1+r_t)\lambda_{t,t+1},$$

where we recall that the interest rate r_t is the interest rate as of period *t* and is assumed by the agent to be equal to all future interest rates, and $\lambda_{t,q}$, $q > t$, is the marginal utility of wealth as of period *q*, (with $\lambda_t = \lambda_{tt}$) given the taste shifters $\{\xi_{\tau 1}, \xi_{\tau 2}, \gamma_{\tau k}, \tau \geq t\}$. From (7.3) it then follows that

(7.7) $$\lambda_t = \lambda_{tT} \beta^{T-t} (1+r_t)^{T-t}.$$

Specifically, the term λ_{tT} will be assumed to have the structure $\lambda_{tT} = \bar{\lambda}_T \kappa_t$, where $\bar{\lambda}_T$ is an individual fixed effect whereas κ_t is a random variable. Recall that in the special case with perfect foresight there are no random variations in tastes as perceived by the agent. This case is captured by letting $\kappa_t = 1$. Define $V_j(t) = \log \tilde{z}_{tj}$, for $j > 0$, and $V_0(t) = \log(-u'_{t2}(0)\xi_{t2} / \lambda_t)$. We see that the agent's behavior at any point in time is completely determined by maximizing the function $V_j(t)$. Assume furthermore that z_{tj} has the structure $z_{tj} = \bar{z}_{tj} \alpha_j \zeta_{tj}$, and $\gamma_{tj} = \bar{\gamma}_j \delta_{tj}$, where $\bar{z}_{tj}$ is a function of observed variables representing the level of human capital, α_j is an individual and sector-specific effect, $\bar{\gamma}_j$ is a sector specific constant and δ_{tj} and ζ_{tj} are random terms. It follows that we can write

(7.8) $$V_j(t) = \log(\bar{z}_{tj}) + \log \alpha_j + \log \bar{\gamma}_j + \log(\delta_{tj} \zeta_{tj}),$$

for $j > 0$, and

(7.9) $$V_0(t) = \log(-u'_{t2}(0)) + \log \bar{\lambda}_T - (T-t)\log(1+r_t) - (T-t)\log(\beta) + \log(\xi_{t2} / \kappa_t).$$

Recall that the terms $\{\delta_{tj} \zeta_{tj}\}$ and $\{\xi_{t2} / \kappa_t\}$ vary over time in a manner that is stochastic to the individual agent due to the fact that she or he is allowed to revise current and future tastes. Now assume that Assumptions 1 to 3 hold. Then it follows from Corollary 1 that the random terms $\{\delta_{tj} \zeta_{tj}\}$ and $\{\xi_{t2} / \kappa_t\}$ can be expressed as $\log(\delta_{tj} \zeta_{tj}) = \eta_j(t) / \sigma$ and $\log(\xi_{t2} / \kappa_t) = \eta_0(t) / \sigma$, where σ is a suitable constant and $\eta(t) = (\eta_0(t), \eta_1(t), ...)$ is standard multivariate type III extreme value distributed at each moment in time, and is also serially independent. The parameter σ is a dispersion parameter that serves to normalize such that $\eta_j(t)$ has standard type III extreme value distribution. Hence it follows from (7.8) and (7.9) that

(7.10) $$U_j(t) = \max(U_j(t-1) - \theta, \sigma V_j(t)).$$

As a result, the choice process, conditional on the fixed effects $\{\alpha_j\}$ and $\bar{\lambda}_T$, is a Markov process with corresponding transition probabilities and probabilities of being in the respective states given by Theorem 2 with $\sigma V_j(t) = w_j(t) + \eta_j(t),$ where

(7.11) $$w_j(t) = \sigma \log\left(\bar{z}_{tj}\right) + \sigma \log(\alpha_j \bar{\gamma}_j),$$

for $j > 0,$ and

(7.10) $$w_0(t) = \sigma \log(-u'_{t2}(0)) + \sigma \log \bar{\lambda}_T - (T-t)\sigma \log(1+r_t) - (T-t)\sigma \log(\beta).$$

Recall that the term $\theta \geq 0,$ is a parameter that represents the degree of habit persistence.

This example, as well as the previous one, demonstrates how one can fruitfully apply the theory developed in this paper to provide theoretically justified functional form and distributional properties of random components in structural models.

8. Conclusion

In this paper we have discussed the issue of functional form of multistate duration models generated by discrete choice over time, with particular reference to random tastes and habit persistence. We have proposed a particular characterization of exogenous random preferences conditional on the information available to the agent. We have demonstrated that this implies preferences that can be represented by modified extremal utility processes. From this result a number of important results follow: First, the choice model under the absence of state dependence and conditional on information available to the agent, is a Markov chain in continuous time. Second, the transition intensities (hazard functions) and probabilities of being in a particular state of this Markov chain model have a tractable structure as functions of the parameters of the utility function. We have furthermore considered extensions that allow for structural state dependence and we have discussed identification and estimation in the presence of fixed and random effects. Finally, we have discussed the application of the framework for modeling occupational mobility in both a static and a life cycle setting without and with uncertainty. The implication of this modeling framework is that provided the researcher knows the distribution of unobservables that are known to the individual agent, one can separate true state dependence effects from habit persistence and unobserved heterogeneity. As we have discussed in this paper, this is particularly true when the unobservables (known to the individual agent), can be represented as fixed or random effects.

References

Chamberlain, G. (1980): Analysis of Covariance with Qualitative Data. *Review of Economic Studies*, **47**, 225-238.

Dagsvik, J. K. (1988): Markov Chains Generated by Maximizing Components of Multi-dimensional Extremal Processes. *Stochastic Processes and Their Applications*, **28**, 31-45.

Dagsvik, J. K. (1994): Discrete and Continuous Choice, Max-stable Processes and Independence from Irrelevant Attributes. *Econometrica*, **62**, 1179-1205.

Dagsvik, J. K. (2002): Discrete Choice in Continuous Time: Implications of an Intertemporal Version of the IIA Property. *Econometrica*, **70**, 817-831.

Fechner, G. T. (1860/1966): *Elemente der Psychophysik*. Druck und Verlag von Breitkopfs Hartel, Leipzig. [*Elements of Psychophysics*, (Vol I). Holt, Rinehart & Winston, 1966, New York].

Gescheider, G. A. (1997): *Psychophysics: The Fundamentals.* Lawrence Erlbaum Associates, Inc., Publishers, New Jersey.

Heckman, J.J. and T. MaCurdy (1980): A Life-Cycle Model of Female Labor Supply. *Review of Economic Studies,* **47**, 47-74.

Heckman, J.J. (1981a): Statistical Models for the Analysis of Discrete Panel Data. In C.F. Manski and D. McFadden (eds.), *Structural Analysis of Discrete Data*. MIT Press, Cambridge, USA.

Heckman, J.J. (1981b): Heterogeneity and State Dependence. In S. Rosen (ed.), *Studies in Labor Markets*. University of Chicago Press, Chicago.

Heckman, J. J., and B. Singer (1984): Econometric Duration Analysis. *Journal of Econometrics*, **24**, 63-132.

Heckman, J.J. (1991): Identifying the Hand of the Past: Distinguishing State Dependence from Heterogeneity. *American Economic Review*, **81**, 75-79.

Herbart, J. F. (1824): *Psychologie als Wissenschaft, neu gegrunndet Auferfarung, Metaphysik, und Matematik.* Unzer, Köningsberg, Germany.

Honoré, B. E. and E. Kyriazidou (2000): Panel Data Discrete Choice Models with Lagged Dependent Variables. *Econometrica*, **68**, 839-874.

Jaggia, S. and P. K. Trivedi (1994): Joint and Separate Score Tests for State Dependence and Unobserved Heterogeneity. *Journal of Econometrics*, **60**, 273-291.

Lee, M.-J. (2002): *Panel Data Econometrics: Methods-of-Moments and Limited Dependent Variables*. Academic Press, San Diego.

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

Luce, R. D. (1959): *Individual Choice Behavior. A Theoretical Analysis*. Wiley, New York.

Luce, R. D. and P. Suppes (1965): Preferences, Utility and Subjective Probability. In R. D. Luce, R. R. Bush and E. Galanter (Eds.), Handbook of mathematical psychology, Vol III, Wiley, New York.

McFadden, D. (1978): Modelling the Choice of Residential Location. In A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull (eds.), *Spatial Interaction Theory and Planning*. North-Holland, Amsterdam.

Magnac, T. (2000): Subsidised Training and Youth Employment: Distinguishing Unobserved Heterogeneity from State Dependence in Labor Market Histories. *Economic Journal*, **110**, 805-837.

Resnick, S. I. (1987): *Extreme Values, Regular Variation and Point Processes*. Springer-Verlag. New York.

Resnick, S. I., and R. Roy (1990): Multivariate Extremal Processes, Leader Processes and Dynamic Choice. *Advances in Applied Probability*, **22**, 309-331.

Rust, J. (1987): Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher. *Econometrica*, **55**, 999-1033.

Strauss, D. (1979): Some Results on Random Utility Models. *Journal of Mathematical Psychology*, **20**, 35-52,

Thurstone, L. L. (1927): A Law of Comparative Judgment. *Psychological Review*, **34**, 273-286.

Tiago de Oliveira, J. (1973): An Extreme Markovian Stationary Process. *Proceedings of the Fourth Conference in Probability Theory* (Brasov, 1971), pp. 217-225. Editura Acad. R. S. R. Bucharest (1973), Romania.

Tversky, A. (1969): Intransitivity of Preferences. *Psychological Review*, **76**, 31-48.

Van den Berg, G. (2001): Duration Model: Specification, Identification and Multiple Durations. In J.J. Heckman and E.E. Leamer (Eds.), *Handbook of Econometrics*, Vol 5, ch. 55, Elsevier, New York.

Appendix A

To prove Theorem 1 we need the following lemmas.

Lemma 1

Let $F(x_1, x_2, ..., x_m)$ *be a type III multivariate extreme value c.d.f. defined on* R^m. *Then* $-\log F(x_1, x_2, ..., x_m)$ *is a convex function.*

Proof of Lemma 1:

From Theorem 5.11´ in Resnick (1987, p. 272) it follows that one can write

$$-\log F(x_1, x_2, ..., x_m) = \int_0^1 \max_{j \le m}(f_j(s)e^{-x_j})ds,$$

where $f_j(s), j = 1, 2, .., m,$ are nonnegative integrable functions such that

$$\int_0^1 f_j(s)ds = 1.$$

Since $f_j(s)e^{-x_j}$ is convex as a function of x_j it follows that $\max_{j \le m}(f_j(s)e^{-x_j})$ is a convex function in $(x_1, x_2, ..., x_m)$ for each *s*. Since sums of convex functions are convex a suitable limit argument yields that the integral of $\max_{j \le m}(f_j(s)e^{-x_j})$ with respect to *s* is convex.

Q.E.D.

Lemma 2

Under Assumtions 1 and 2 the distribution of $\varepsilon(t) = U(t) - v(t)$ *is a multivariate extreme value distribution type III which is independent of time.*

Proof of Lemma 2:

From Assumption 2 it follows that at each given point in time *t*, $\varepsilon(t)$ is independent of $v(t)$, where can vary freely on the real line. It will be convenient to normalize such that $E\varepsilon_j(t) = 0.5772$ (Euler's constant). Under Assumptions 1 and 2 it follows from Lindberg et al. (1995) that the joint distribution of $\varepsilon(t)$ has the form $\psi(\exp(-\tilde{H}(e^{-y_1}, e^{-y_2}, ..., e^{-y_m}))),$ where ψ a positive and strictly increasing mapping on $[0,1]$ with $\psi(0) = 0$ and $\psi(1) = 1,$ and $\tilde{H}(x_1, x_2, ..., x_m)$ is a positive increasing and linear homogenous function defined on $[0, \infty)^m$, and may depend on time. In addition, $\tilde{H}$ must satisfy suitable regularity conditions that ensures that $\psi(\exp(-\tilde{H}(e^{-y_1}, e^{-y_2}, ..., e^{-y_m})))$ is a multivariate c.d.f.

(McFadden, 1978). Since by Assumption 2, $\varepsilon_1(t)$ is independent of $\varepsilon_j(t)$ for $j \in S \setminus \{1\}$, the joint distribution of $(\varepsilon_1(t), \varepsilon_j(t))$ must be a product of type III extreme value distributions, which implies that

$$P(\max(\varepsilon_1(t), \varepsilon_j(t)) \le y) = \exp(-e^{-y} - e^{-y}) = \psi(\exp(-\tilde{H}(e^{-y}, 1, 1, .., e^{-y}, 1, ., 1)))$$
$$= \psi(\exp(-e^{-y}\tilde{H}(1, 1, ..., 1))).$$

The last equation implies that $\psi(x) = x^a$, for $x \in [0,1]$, where a is an arbitrary positive constant. Hence, we have proved that the c.d.f of $\varepsilon(t)$ is a type III multivariate extreme value distribution at each given point in time.

Q.E.D.

Let *A* be any subset of *S*. Define $U_A(t) = \max_{k \in A} U_k(t)$.

Lemma 3

Assume that Axioms 1 and 2 hold. Then the joint distribution of $(U_A(s), U_A(t))$, *for s < t, is a bivariate extreme value distribution (type III).*

Proof of Lemma 3:

Let the function $G_A(u_1, u_2)$ be defined on $[0, \infty) \times [0, \infty)$ by

(A.1) $$\exp(-G_A(e^{-x}, e^{-y})) = P\left(U_A(s) \le x, U_A(t) \le y\right),$$

for $A \subset S$. For notational convenience we normalize such that $G_1(1,0) = G_A(1,0) = 1$. This normalization is possible and does not imply any loss of generality since, by Assumption 2, the structural part of the utility function is independent of the random part. Let *z* be an arbitrary real number. Let now $A \subset S \setminus \{1\}$ and note that the process $\{(U_1(t), U_A(t) + z), t \ge 0\}$ also satisfies Assumptions 1 to 3, provided the process $\{(U_1(t), U_A(t)), t \ge 0\}$ satisfies Assumptions 1 to 3. Lemma 2 implies that $U_A(t)$ is type III extreme value distributed. Since the utility processes associated with state 1 and *j* are independent, we obtain that

(A.2)
$$\begin{aligned}
&P\left(U_A(s) + z < U_1(s), \max\left(U_1(s), U_A(s) + z\right) \in (x, x + dx), \max\left(U_1(t), U_A(t) + z\right) \le y\right) \\
&= P\left(U_A(s) + z < U_1(s) \in (x, x + dx), \max\left(U_1(t), U_A(t) + z\right) \le y\right) \\
&= P\left(U_A(s) \le x - z, U_A(t) \le y - z, U_1(s) \in (x, x + dx), U_1(t) \le y\right) \\
&= P\left(U_A(s) \le x - z, U_A(t) \le y - z\right) P\left(U_1(s) \in (x, x + dx), U_1(t) \le y\right) \\
&= \exp\left(-G_A(e^{z-x}, e^{z-y}) - G_1(e^{-x}, e^{-y})\right) e^{-x} \partial_1 G_1(e^{-x}, e^{-y})) dx,
\end{aligned}$$

where ∂_k denotes the partial derivative with respect to the k-th component. Similarly

$$P\left(\max\left(U_1(s),U_A(s)+z\right)\le x,\max\left(U_1(t),U_A(t)+z\right)\le y\right)=\exp\left(-G_1(e^{-x},e^{-y})-G_A\left(e^{z-x},e^{z-y}\right)\right),$$

which implies that

(A.3)
$$\begin{aligned}&P\left(\max\left(U_1(s),U_A(s)+z\right)\in(x,x+dx),\max\left(U_1(t),U_A(t)+z\right)\le y\right)\\&=\exp\left(-G_1(e^{-x},e^{-y})-G_A\left(e^{z-x},e^{z-y}\right)\right)\left(\partial_1G_1(e^{-x},e^{-y})+e^z\partial_1G_A\left(e^{z-x},e^{z-y}\right)\right)e^{-x}dx.\end{aligned}$$

Assumption 1 implies that

(A.4)
$$\begin{aligned}&P\left(\max\left(U_1(s),U_A(s)+z\right)\in\left(x,x+dx\right),\max\left(U_1(t),U_A(t)+z\right)\le y,J_B(s)=1\right)\\&=P\left(\max\left(U_1(s),U_A(s)+z\right)\in\left(x,x+dx\right),\max\left(U_1(t),U_A(t)+z\right)\le y\right)P\left(U_1(s)>U_A(s)+z\right).\end{aligned}$$

When (A.1) and (A.3) are inserted into (A.4) we obtain

$$\begin{aligned}&\exp\left(-G_1\left(e^{-x},e^{-y}\right)-G_A\left(e^{z-x},e^{z-y}\right)\right)\partial_1G_1\left(e^{-x},e^{-y}\right)\\&=P(U_1(s)>U_A(s)+z)\cdot\exp\left(-G_1\left(e^{-x},e^{-y}\right)-G_A\left(e^{z-x},e^{z-y}\right)\right)\left(\partial_1G_1\left(e^{-x},e^{-y}\right)+e^z\partial_1G_A\left(e^{z-x},e^{z-y}\right)\right)\end{aligned}$$

which is equivalent to

(A.5)
$$\partial_1G_1\left(e^{-x},e^{-y}\right)P(U_1(s)<U_A(s)+z)=e^z\partial_1G_A\left(e^{z-x},e^{z-y}\right)P(U_1(s)>U_A(s)+z).$$

Furthermore, since $U_1(s)$ and $U_A(s)$ are independent, $G_A(1,0)=G_1(1,0)=1$, $G_A(e^z,0)=e^zG_A(1,0)$, it follows that

(A.6)
$$P(U_1(s)<U_A(s)+z)=\frac{G_A(e^z,0)}{G_1(1,0)+G_A(e^z,0)}=\frac{G_A(1,0)e^z}{G_1(1,0)+G_A(1,0)e^z}=\frac{e^z}{1+e^z}.$$

When (A.6) is inserted into (A.5) it follows that

(A.7)
$$\partial_1G_1\left(e^{-x},e^{-y}\right)=\partial_1G_A\left(e^{z-x},e^{z-y}\right).$$

By integrating (A.7) with respect to *x* we obtain

(A.8)
$$G_1\left(e^{-x},e^{-y}\right)=e^{-z}G_A\left(e^{z-x},e^{z-y}\right)+K_A(y),$$

for some function $K_A(y)$ that does not depend on *x* nor *z*. Since $G_1(0,e^{-y})=G_1(0,1)e^{-y}$ and $e^{-z}G_A(0,e^{z-y})=G_A(0,e^{-y})=G_A(0,1)e^{-y}$, it follows from (A.8), with $z = 0$, that

(A.9)
$$K_A(y)=(G_1(0,1)-G_A(0,1))e^{-y}.$$

From (A.8) and (A.9) we therefore get that

(A.10)
$$e^{-z}G_A\left(e^{z-x},e^{z-y}\right)=G_1\left(e^{-x},e^{-y}\right)+(G_A(0,1)-G_1(0,1))e^{-y}.$$

Eq. (A.10) is valid for any real *z* and since the right hand side of (A.10) does not depend on *z* we get

(A.11)
$$e^{-z}G_A\left(e^{z-x},e^{z-y}\right)=G_A\left(e^{-x},e^{-y}\right),$$

for all real *z*. Similarly, it follows from (A.10) and (A.11) that

$$e^{-z}G_1(e^{z-x},e^{z-y}) = (G_1(0,1) - G_A(0,1))e^{-y} + e^{-z}G_A(e^{z-x},e^{z-y})$$

$$= (G_1(\infty,0) - G_A(\infty,0))e^{-y} + G_A(x,y) = G_1(x,y)\,,$$

which proves that (A.11) also holds for A = {1}. But this means that the bivariate c.d.f. defined in (A.1) is a bivariate type III extreme value distribution for any $A \in S$, and therefore the proof of Lemma 3 is complete.

Q.E.D.

Proof of Theorem 1:

For $A \subset S$, let $\varphi_A(z) = G_A(e^z,1)$ where G_A is defined in (A.1). For $A \subset S\setminus\{1\}$ define $\psi_A(z) = \varphi_1(z) + \varphi_A(z).$ From these definitions it follows that for any $A \subset S$,

(A.12) $$G_A(e^{-x},e^{-y}) = e^{-y}G_A(e^{y-x},1) = e^{-y}\varphi_A(y-x)$$

and

(A.13) $$e^{-x}\partial_1 G_A(e^{-x},e^{-y}) = e^{-y}\varphi_A'(y-x) \text{ and } e^{-y}\partial_2 G_A(e^{-x},e^{-y}) = e^{-y}(\varphi_A(y-x) - \varphi_A'(y-x)).$$

Moreover, for $A \in S\setminus\{1\}$ and $B = \{1, A\}$,

(A.14) $$P\left(\max_{k\in B} U_k(t) \le y\right) = \exp(-G_1(0,e^{-y}) - G_A(0,e^{-y})) = \exp(-e^{-y}\psi_A(-\infty)).$$

Let

$$C_{ij} = P(J_B(s) = i, J_B(t) = j)\psi_A(-\infty),$$

for $i, j \in B.$ From (A.12) to (A.14) and Assumption 1 we get that

$$C_{1A}\exp(-e^{-y}\psi(-\infty))e^{-y}\Delta y + o(\Delta y) = P(J_B(s) = 1, J_B(t) \in A, U_B(t) \in (y, y+\Delta y))$$

$$= \int_R P(U_A(s) \le x, U_1(s) \in (x, x+dx), U_1(t) \le y, U_A(t) \in (y, y+\Delta y))$$

$$= \Delta y \int_R \partial_1 \exp(-G_1(e^{-x},e^{-y}) - G_A(e^{-x},e^{-y}))\partial_1 G_1(e^{-x},e^{-y})\partial_2 G_A(e^{-x},e^{-y})e^{-x-y}dx$$

$$= \Delta y \int_R \exp(-e^{-y}\psi_A(y-x))e^{-2y}\varphi_1'(y-x)(\varphi_A(y-x) - \varphi_A'(y-x))dx.$$

By dividing the last equation by $e^{-y}\Delta y$ and letting $\Delta y \to 0,$ we obtain that

(A.15) $$C_{1A}\exp(-e^{-y}\psi(-\infty)) = e^{-y}\int_R \exp(-e^{-y}\psi_A(y-x))\varphi_1'(y-x)(\varphi_A(y-x) - \varphi_A'(y-x))dx.$$

Note furthermore that since $\psi_A(\infty) = \infty,$ one can write

(A.16) $$\exp(-e^{-y}\psi_A(-\infty)) = e^{-y}\int_{-\infty}^{\infty} \exp(-e^{-y}\psi_A(z))\psi_A'(z)dz.$$

By change-of-variable; $z = y - x$ in the integral in (A.15), (A.15) and (A.16) yield

(A.17) $$C_{1A}\int_{-\infty}^{\infty}\exp(-e^{-y}\psi(z))\psi'(z)dz=\int_{-\infty}^{\infty}\exp(-e^{-y}\psi(z))\varphi_1'(z)(\varphi_A(z)-\varphi_A'(z))dz.$$

Note next that by Lemma 1, $\varphi_1(z)$, $\varphi_A(z)$ and $\psi_A(z)$ are increasing and convex functions, which implies that $\psi_A'(z)$ is non-decreasing. Define $c=-\infty$ if $\psi_A'(z)>0$ for all real z, otherwise define $c=\max\{z:\psi_A'(z)=0\}$. It thus follows that ψ_A is strictly increasing and invertible for $z > c$. Let $\Omega=\{w:w=\psi_A(z),z>0\}$. Let h be the inverse function of ψ_A, defined on Ω. By the change of variable, $w=h(z)$, (A.17), with $\lambda=e^{-y}$, becomes

(A.18) $$C_{1A}\int_{\Omega}\exp(-\lambda w)dw=\int_{\Omega}\exp(-\lambda w)\varphi_1'(h(w))(\varphi_A(h(w))-\varphi_A'(h(w)))\frac{dw}{\psi_A'(h(w))},$$

which must hold for every positive λ. By the uniqueness property of the Laplace transform (A.18) implies that

$$C_{1A}=\frac{\varphi_1'(h(w))(\varphi_A(h(w))-\varphi_A'(h(w)))}{\psi_A'(h(w))},$$

for $w\in\Omega$, which is equivalent to

(A.19) $$C_{1A}\psi_A'(z)=\varphi_1'(z)(\varphi_A(z)-\varphi_A'(z)).$$

for $z>c$. Similarly, it follows by symmetry that

(A.20) $$C_{A1}\psi_A'(z)=\varphi_A'(z)(\varphi_1(z)-\varphi_1'(z)),$$

for $z>c$. By subtracting (A.20) from (A.19) and dividing by $\psi(z)^2$ we obtain

(A.21) $$\frac{\varphi_1'(z)\psi_A(z)-\varphi_1(z)\psi_A'(z)}{\psi_A(z)^2}=\frac{\psi_A'(z)(C_{1A}-C_{A1})}{\psi_A(z)^2}.$$

When we integrate both sides of (A.21) we get that

$$\frac{\varphi_1(z)}{\psi_A(z)}=\frac{C_{1A}-C_{A1}}{\psi_A(z)}+d_1,$$

for $z>c$, where d_1 is an arbitrary constant. This is equivalent to

(A.22) $$\varphi_1(z)=C_{1A}-C_{A1}+d_1\psi_A(z).$$

Similarly, it follows by symmetry that

(A.23) $$\varphi_A(z)=C_{A1}-C_{1A}+d_A\psi_A(z),$$

where d_A is an arbitrary constant. When the equation in (A.22) is inserted for $\varphi_1(z)$ into eq. (A.19) it follows that

(A.24) $$\varphi_A(z)-\varphi_A'(z)=\frac{C_{1A}}{d_1}.$$

Similarly, when the equation in (A.23) is inserted for $\varphi_A(z)$ into eq. (A.20) it follows that

(A.25) $$\varphi_1(z) - \varphi_1'(z) = \frac{C_{A1}}{d_A}.$$

Eq. (A.24) and (A.25) are first order differential equations that have solutions of the form

(A.26) $$\varphi_j(z) = \alpha_j + \beta_j e^z,$$

for j = 1 and j = A, respectively, and $z > c$, where α_j and β_j are suitable constants. Since $\varphi_j'(z) = 0$ for $z \le c$ and $\varphi_j(z)$ is continuous it follows from (A.26) that we can write

(A.27) $$\varphi_j(z) = \alpha_j + \beta_j \exp(\max(c, z)),$$

for any real z. Using (A.12) we obtain, as a consequence of (A.27), that

(A.28) $$G_j(e^{-x}, e^{-y}) = \alpha_j e^{-y} + \beta_j e^{-y} \exp(\max(c, y - x)) = \alpha_j e^{-y} + \beta_j \exp(-\min(x, y - c)),$$

for j = 1, A. The equation in (A.28) implies that the utility process $\{U_j(t), t \ge 0\}$, $j \in B$, is equivalent to the utility process generated by

(A.29) $$U_j(t) = \max(U_j(s) + c, W_j(s,t)),$$

for $j = 1, A$, $A \subset S \setminus \{1\}$, where $W_j(s,t)$ is type III extreme value distributed and independent of $U_j(s)$. Since A can be a singleton, we conclude that (A.29) must hold for any $j \in S$.

We next prove that $c = -(s - t)\theta$, where θ is a positive constant. To emphasize that c depends on s and t we write henceforth $c = c(s, t)$. Under Assumption 3 it must be true that $c(s,t) = c(t - s)$. From (A.29) it follows for $s < r < t$, that

$$U_j(t) = \max(U_j(r) + c(r,t), W_j(r,t)) \text{ and } U_j(r) = \max(U_j(s) + c(s,r), W_j(s,r)),$$

which, together with (A.31) imply that

(A.30) $$U_j(t) = \max(U_j(s) + c(s,r) + c(r,t), W_j(s,r) + c(r,t), W_j(r,t)) = \max(U_j(s) + c(s,t), W_j(s,t)).$$

Furthermore, (A,29) and (A.30) imply that $c(t - s) = c(r - s) + c(t - r)$, which must hold for any $s < r < t$. Let $x = r - s$ and $y = t - r$. The latter equation is equivalent to

(A.31) $$c(x + y) = c(x) + c(y),$$

which must hold for all positive x and y. The equation in (A.31) is a functional equation of the Cauchy type and the only possible solution is $c(x) = -x\theta$, where θ is a constant, see for example Theorem 3.2 in Falmagne (1985, p. 82). The constant θ must be positive because otherwise $\{U_j(t), t \ge 0\}$ cannot be stationary, as is easily verified. Moreover, Tiago de Oliveira (1973) has demonstrated that the representation (2.1) in fact is a strictly stationary (Markovian) stochastic process in continuous time.

Finally, we shall prove that the corresponding vector equation $U(t) = \max(U(s) + c, W(s,t))$ holds, with the vector $W(s,t)$ independent of $U(s)$. Without essential loss of generality and for the sake of simplicity we shall only give the proof for the bivariate case, that is, for the case where $U(s)=(U_2(s),U_3(s))$ and $W(s,t)=(W_2(s,t),W_3(s,t)).$ Recall that we have proved that $U_j(s)$ is independent of $W_j(s,t)$ for any *j*, but it remains to prove that $W_2(s,t)$ is independent of $U_3(s),$ and $W_3(s,t)$ is independent of $U_2(t).$ Recall that we have demonstrated that $U_A(s)$ and $W_A(s,t)$ are independent and type III extreme value distributed for any $A\subset S.$ Let $\varepsilon_j(t)=U_j(t)-v_j(t)$ and $\xi_j(s,t)=W_j(s,t)-w_j(s,t),$ where $\varepsilon_j(s)$ and $\xi_j(s,t)$ are standard type III extreme value distributed variables, and $v_j(t)$ and $w_j(s,t)$ are deterministic terms. It follows from (A.29) that $w_j(s,t)=\log(\exp(v_j(t))-\exp(v_j(s)-(t-s)\theta)).$ By Lemma 2 we can write

(A.32) $$P(\varepsilon_2(t)\le x,\varepsilon_3(t)\le y)=\exp(-H_{23}(e^{-x},e^{-y}))$$

where the function $H_{23}(u,v)$ is defined on $[0,\infty)\times[0,\infty)$ and is positive, linear homogeneous and increasing, and is independent of time. Because of (A.29) we have, with A = {2,3}, that

(A.33) $$\max(\varepsilon_2(s)+v_2(s)-(t-s)\theta,\varepsilon_3(s)+v_3(s)-(t-s)\theta,\xi_2(s,t)+w_2(s,t),\xi_3(s,t)+w_3(s,t))$$
$$=\max(\varepsilon_2(t)+v_2(t),\varepsilon_3(t)+v_3(t)).$$

Consider the special case with $v_j(s)=v_j(t)=v_j,$ implying that $w_j(s,t)=v_j+\log(1-\exp(-(t-s)\theta))\equiv v_j+\kappa.$ Then, in this case (A.33) implies that

(A.34) $$P(\varepsilon_2(s)\le y+(t-s)\theta-v_2,\varepsilon_3(s)\le y+(t-s)\theta-v_3)P(\xi_2(s,t)\le y-v_2-\kappa,\xi_3(s,t)\le y-v_3-\kappa)$$
$$=P(\varepsilon_2(t)\le y-v_2,\varepsilon_3(t)\le y-v_3).$$

If we combine (A.32) with (A.34) we obtain, by using the linear homogeneity property of $H_{23},$ that

$$P(\xi_2(s,t)\le y-v_2-\kappa,\xi_3(s,t)\le y-v_3-\kappa)=\exp(-H_{23}(e^{v_2-y},e^{v_3-y})+H_{23}(e^{v_2-y-(t-s)\theta},e^{v_3-y-(t-s)\theta}))$$
$$=\exp(-H_{23}(e^{v_2-y},e^{v_3-y})+e^{-(t-s)\theta}H_{23}(e^{v_2-y},e^{v_3-y}))$$
$$=\exp(-e^{\kappa}H_{23}(e^{v_2-y},e^{v_3-y}))=\exp(-H_{23}(e^{\kappa+v_2-y},e^{\kappa+v_3-y})).$$

Since $v_j, j=2,3,$ can vary freely the latter equation is equivalent to

(A.35) $$P(\xi_2(s,t)\le y_2,\xi_3(s,t)\le y_3)=\exp(-H_{23}(e^{-y_2},e^{-y_3})),$$

where $y_j, j=2,3,$ are any real numbers. Hence we have proved that the c.d.f. of $(\xi_2(s,t),\xi_3(s,t))$ is independent of time and of the structural terms $w_j(s,t), j=2,3.$ Furthermore, since $U_A(s)$ and $W_A(s,t)$ are independent we have that

$$P(U_A(s)\le x,W_A(s,t)\le y)=P(U_A(s)\le x)P(W_A(s,t)\le y)$$
$$=P(\varepsilon_2(s)\le x-v_2(s),\varepsilon_3(s)\le x-v_3(s))P(\xi_2(s,t)\le y-w_2(s,t),\xi_3(s,t)\le y-w_3(s,t)).$$

Note that the terms $v_j(s)$ and $w_j(s,t), j=2,3,$ can vary freely on the real line. Thus the last equation is equivalent to

(A.36) $$P(\varepsilon_2(s)\le x_2,\varepsilon_3(s)\le x_3,\xi_2(s,t)\le y_2,\xi_3(s,t)\le y_3)$$
$$=P(\varepsilon_2(s)\le x_2,\varepsilon_3(s)\le x_3)P(\xi_2(s,t)\le y_2,\xi_3(s,t)\le y_3).$$

Eq. (A.36) implies that $(\varepsilon_2(s),\varepsilon_3(s))$ and $(\xi_2(s,t),\xi_3(s,t))$ are independent, which implies that $(U_2(s),U_3(s))$ and $(W_2(s,t),W_3(s,t))$ are independent. This completes the proof.

Q.E.D.

Proof of Theorem 2:

We shall demonstrate Theorem 2 follows from Dagsvik (1988). Let $G_t(x)$ be defined as in Dagsvik (1988), namely by $G_t(x)=-\log P(U(t)\le x)=-\log P(W(0,t)\le x).$ Then it follows that

$$V_j(t)=-\frac{\partial G_t(x)}{\partial x_j}\Big|_{x=0}.$$

Furthermore, note that

$$\sum_{k\in B}V_k(t)=G_t(0).$$

Now the result of Theorem 2 follows from Theorem 3, p. 34, of Dagsvik (1988).

Q.E.D.

Lemma 4

Let $\{Z(t),t\ge 0\}$ be a standard extremal process defined as follows:

(i) $Z(0)=-\infty,$

(ii) $Z(t)=\max(Z(s),Y(s,t)),$ *for s < t, where* $Y(s,t)$ *is a random variable that is independent of* $Z(s)$,

(iii) $Y(s,t)$ *and* $Y(s',t')$ *are independent whenever* $(s,t)\cap(s',t')=\varnothing$,

(vi) $P(Y(s,t)\le u)=\exp(-(t-s)e^{-u}),$ *for real u.*

If $U_j(t), j\in S,$ *are independent modified extremal utility processes (as defined in Theorem 1) we have that*

(A.37) $$U_j(t)=Z_j(V_j(t))-\theta t,$$

where $\{Z_j(t),t\ge 0\}, j\in S,$ *are independent standard extremal processes and*

(A.38) $$V_j(t)=\int_0^t\exp(w_j(\tau)+\theta\tau)d\tau+c_j,$$

where $\theta \geq 0$ *is a constant and* $\{w_j(\tau)\}$ *are deterministic terms.*

Proof of Lemma 4:

Consider the three-dimensional distributions of $\{Z(t), t \geq 0\}$. It follows from the definition above in the statement of the lemma that for real x_j, j = 1, 2, 3, and $t_1 < t_2 < t_3$, that

$$P(Z(t_1) \leq x_1, Z(t_2) \leq x_2, Z(t_3) \leq x_3) = \exp(-t_1 e^{-\min(x_1,x_2,x_3)} - (t_2 - t_1)e^{-\min(x_2,x_3)} - (t_3 - t_2)e^{-x_3}).$$

Furthermore, it follows from the definition of H in (2.2) that when the utility processes are independent that

$$H(y_1, y_2, ..., y_m) = y_1 + y_2 + \cdots + y_m,$$

so that

$$P(W_j(s,t) + \theta t \leq u) = \exp\left(-e^{-u}\int_s^t \exp(w_j(\tau) + \tau\theta)d\tau\right) = \exp(-e^{-u}(V_j(t) - V_j(s))).$$

Let $\hat{U}_j(t) = U_j(t) + \theta t.$ The corresponding updating relation is given by

$$\hat{U}_j(t) = \max(\hat{U}_j(s), W_j(s,t) + \theta t).$$

Moreover, (A.37) is equivalent to $\hat{U}_j(t) = Z_j(V_j(t)).$ Hence, we obtain that

$$P(U_j(t_1) \leq x_1, U_j(t_2) \leq x_2, U_j(t_3) \leq x_3)$$

$$P(\hat{U}_j(t_1) \leq x_1, \max(\hat{U}_j(t_1), W_j(t_1,t_2) + \theta t_2) \leq x_2, \max(\hat{U}_j(t_1), W_j(t_1,t_2) + \theta t_2, W_j(t_2,t_3) + \theta t_3)) \leq x_3)$$

$$= \exp(-V_j(t_1)e^{-\min(x_1,x_2,x_3)} - (V_j(t_2) - V_j(t_1))e^{-\min(x_2,x_3)} - (V_j(t_3) - V_j(t_2))e^{-x_3}).$$

It is easy to verify that the last expression is same as

$$P(Z_j(V_j(t_1)) \leq x_1, Z_j(V_j(t_2)) \leq x_2, Z_j(V_j(t_3)) \leq x_3).$$

For the higher dimensional distributions the computation is entirely analogous. This completes the proof.

Q.E.D.

Proof of Theorem 3:

For any choice set B, let $I_j(B) = 1$ if $j \in B,$ and zero otherwise. Define independent utility processes, $\{U_j^*(t), t \geq 0\}$, by

(A.39) $$U_j^*(t) = Z_j(V_j^*(t)) - \theta t,$$

where $\{Z_j(t), t \geq 0\}$, j =1, 2, …., *m*, are independent standard extremal processes, and $V_j^*(t) = V_j(t) I_j(B(t)).$ Evidently, $V_j^*(t)$ is a non-decreasing function in *t* when the choice set process is non-decreasing. Consequently, Lemma 4 implies that the utility processes defined in (A.39) are modified extremal processes. Thus, the results of Theorem 2 and Corollary 2 go through with $V_j(t)$ replaced by $V_j^*(t).$ Hence, the result of Theorem 3 follows.

Q.E.D.

Proof of Theorem 4:

Let $U_j(t, h(t))$ be defined as in Assumption 4. Let $U(t, h(t)) = (U_1(t, h(t)), U_2(t, h(t)))$ and $w(t, h(t)) = (w_1(t, h(t)), w_2(t, h(t))).$ For notational simplicity, define

(A.40) $$M_t\left(y \middle| x, h(t)\right) = \begin{cases} \exp\left(-\exp\left(w_1\left(t, h(t)\right) - y_1\right) - \exp\left(w_2\left(t, h(t)\right) - y_2\right)\right) & \text{for } \; y_j \geq x_j + \theta, j = 1, 2, \\ 0 \quad \text{otherwise}. \end{cases}$$

Due to the form of the extreme value distribution it follows that the conditional distribution of $U(t, h(t))$ given $U(t-1, h(t-1))$ and given the choice history can be expressed as

(A.41) $$P\left(U(t, h(t)) \leq y \middle| U(t-1, h(t-1)) = x\right) = M_t\left(y \middle| x, h(t)\right) \text{ for } \quad y_j \geq x_j + \theta, j = 1, 2,$$

and zero otherwise. Let $j(\tau), \tau = 1, 2, ...,$ be a sequence of choices and define $u(t) = (u_1(t), u_2(t))$ and

$$\Omega(t) = \left\{\left(u(1), u(2), ..., u(t)\right) : u_{j(\tau)}(\tau) = \max_k u_k(\tau), \forall \tau \leq t\right\}.$$

We have

(A.42) $$P\left(J(\tau) = j(\tau), \tau = 1, 2, ..., t\right) = \int_{\Omega(t)} \prod_{\tau=1}^{t} d\, M_\tau\left(u(\tau) \middle| u(\tau-1), h^*(\tau)\right)$$

where $u(0) = (-\infty, -\infty)$, and

$$h^*(\tau) = \left\{J(s) = j(s), \forall s \leq \tau - 1\right\}.$$

Let $M_t(y|x)$ be given by (A.40) when $w\left(t, h(t)\right)$ is replaced by $w(t)$. Define

(A.43) $$L\left(h^*(t+1), w(s), s \leq t\right) \equiv \int_{\Omega(t)} \prod_{\tau=1}^{t} dM_\tau\left(u(\tau) \middle| u(\tau-1)\right).$$

Clearly, (A.43) is the likelihood function under pure taste persistence. We know from Theorem 2 that the likelihood function in (A.43) has a structure that implies a Markovian choice process under pure taste persistence. But when $\left(U(1), U(2), ..., U(t)\right) \in \Omega(t)$, then the choice history, including the choice at time t, equals $h^*(t+1)$ and consequently the functions, $w\big(s,h(s)\big)$, $s=1,2,...,t,$ remain fixed when the integrand in (A.43) is integrated out over $\Omega(t)$. Therefore, we must have that

(A.44) $$P\big(J(\tau)=j(\tau),\,\tau=1,2,...,t\big)=L\Big(h^*(t+1),\tilde{w}(s,h^*(s)),s\le t\Big).$$

Eq. (A.44) implies that state dependence can be conveniently accounted for by replacing $\{w(s)\}$ by corresponding "experience-modified" functions $\{w(s,h^*(s))\}$. Furthermore, conditional on these experience-modified functions the choice process is a Markov chain with transition probabilities that have the same structure as in Theorem 2 with $\{w(s),\,s\le t\}$ replaced by $\{w(s,h^*(s)),s\le t\}$.

Q.E.D.

Proof of Corollary 8:

From (3.10) it follows that

(A.45)
$$\int_d^t\left(\sum_{k\in A}q_{ik}(x)+\sum_{k\in B\setminus A}q_{jk}(x)\right)dx=\int_d^t\frac{\sum_{k\in B}V_k'(x)dx}{\sum_{k\in B}V_k(x)}$$
$$=\log\left(\sum_{k\in B}V_k(t)\right)+\kappa=\log\left(\int_0^t e^{\theta\tau}H^B\big(\exp(w(\tau))\big)d\tau+\sum_{k\in B}c_k\right)+\kappa,$$

which proves (6.1). Also from (3.10) we have that

$$\log V_j'(t)=\log q_{ij}(t)+\log\left(\sum_{k\in B}V_k(t)\right).$$

When the above equation is combined with (A.46) we get that

$$\log H_j^B\big(\exp(w(t))\big)+w_j(t)+\theta t=\log q_{ij}(t)+\int_d^t\left(\sum_{k\in A}q_{ik}(x)+\sum_{k\in B\setminus A}q_{jk}(x)\right)dx,$$

which proves (6.2)

Q.E.D.

Appendix B

In this appendix we outline how the transition probabilities and intensities look like when the functions $w_j(t)$ are step functions that only may change at given time units, such as weeks or months. In this case we obtain that

(B.1)
$$\int_s^t \exp\left(w_j(\tau)-(t-\tau)\theta\right)d\tau = \sum_{r=s}^{t}\exp\left(w_j(r)-t\theta\right)\int_{r-1}^{r} e^{\theta\tau}d\tau = \sum_{r=s}^{t}\exp\left(w_j(r)-t\theta\right)\left(e^{\theta r}-e^{\theta r-\theta}\right)\theta^{-1}$$
$$=\frac{\left(1-e^{-\theta}\right)}{\theta}\sum_{r=s}^{t}\exp\left(w_j(r)-(t-r)\theta\right).$$

As a result of (B.1) it follows from Theorem 2 that the transition and state probabilities can be written as

(B.2)
$$Q_{ij}(s,t)=\frac{\sum_{r=s}^{t}\exp\left(w_j(\tau)-(t-\tau)\theta\right)}{\sum_{k=0}^{1}\left(\sum_{r=1}^{t}\exp\left(w_k(\tau)-(t-\tau)\theta\right)d\tau+(\tilde{c}_k+\exp(w_k(0)))e^{-t\theta}\right)}$$

for $0<s\le t, i, j=0,1, i\neq j$, and

(B.3)
$$P_j(t)=\frac{\sum_{r=1}^{t}\exp\left(w_j(\tau)-(t-\tau)\theta\right)d\tau+(\tilde{c}_j+\exp(w_j(0)))e^{-\theta t}}{\sum_{k=0}^{1}\left(\sum_{r=1}^{t}\exp\left(w_k(\tau)-(t-\tau)\theta\right)d\tau+(\tilde{c}_k+\exp(w_k(0)))e^{-\theta t}\right)},$$

where

$$\tilde{c}_j=\frac{\theta c_j}{1-e^{-\theta}}.$$

The corresponding transition intensities follow from (3.15) and (B.1) and are given by

(B.4)
$$q_{ij}(t)=\frac{\theta\exp(w_j(t))}{\sum_{k=0}^{1}\left(\sum_{r=1}^{t}\exp\left(w_k(\tau)-(t-\tau)\theta\right)d\tau+(\tilde{c}_k+\exp(w_k(0)))e^{-\theta t}\right)}.$$

From (B.2), (B.3) and (B.4) it is evident that $\tilde{c}_j$ may be absorbed in $\exp(w_j(0))$, and we can therefore normalize by setting $\tilde{c}_j=0$.

B

Returadresse:
Statistisk sentralbyrå
NO-2225 Kongsvinger

Statistics Norway

Oslo:
PO Box 8131 Dept
NO-0033 Oslo
Telephone: + 47 21 09 00 00
Telefax: + 47 21 09 00 40

Kongsvinger:
NO-2225 Kongsvinger
Telephone: + 47 62 88 50 00
Telefax: + 47 62 88 50 30

E-mail: ssb@ssb.no
Internet: www.ssb.no

ISSN 0809-733X

Statistisk sentralbyrå
Statistics Norway