On the Structure of Behavioral Multistate Duration Models

by

John K. Dagsvik

Statistics Norway

and

Department of Economics,

University of Oslo

Abstract

This paper proposes a particular behavioral assumption to characterize the stochastic structure of intertemporal discrete choice models in the absence of state dependence. This assumption extends Luce's axiom; "Independence from Irrelevant Alternatives", to the intertemporal context. Under certain regularity conditions the implication of this assumption is that the individual choice process is a Markov chain with transition probabilities that have a particularly simple structure. It is demonstrated that this structure is consistent with an intertemporal and life cycle consistent random utility model where the utilities are independent extremal processes in time. Finally, the framework is extended to allow for state dependence and certain types of time-varying choice sets.

Keywords: Life cycle consistent discrete choice, taste persistence, state dependence, Markovian choice processes, extremal processes, random utility processes, independence from irrelevant alternatives.

Acknowledgement: Thanks to Rolf Aaberge, Leif Andreassen, Tor Jakob Klette, Tom Kornstad, Pedro de Lima, Chris Flinn and Steinar Strøm for criticism and comments and Anne Skoglund for technical assistance and correction of errors. I also thank participants in workshop at the University of California, Berkeley, University of Chicago and University of Wisconsin for comments on earlier versions of the paper.

1. Introduction

This paper discusses choice of functional form in structural models for duration analysis. In the context of econometric modelling of duration data it is, for example, often convenient to apply an econometric framework based on the proportional hazard rate formulation. Although the proportional hazard rate formulation is mathematically convenient, it is nevertheless quite insatisfactory because it is ad hoc from a theoretical point of view. A theoretical justification supporting functional form and distributional assumptions is important for the issue of identifying structural effects and for making inferences about the nature and significance of such effects. To highlight the importance of this issue, consider the well-known problem of separating structural state dependence effects from effects due to serially correlated tastes. These effects may arise when unobservables are correlated over time; in which case current and past choices will be correlated. This identification problem is crucial in a variety of contexts and has been discussed most notably by Heckman (see Heckman, 1978, 1981a, 1981b, 1991, and the references therein). For example, in analyses of unemployment it is often noted that individuals who have experienced unemployment in the past are more likely to experience it in the future than are individuals who have not been unemployed. There may be two explanations for this empirical regularity. One explanation is that current and past choices are correlated due to serially correlated unobservables that affect preferences (pure taste persistence) or opportunities. In this case, past choices are proxies for unobserved variables that affect preference evaluations or unobservable opportunity sets, and consequently the (aggregate) transition rates will depend on past choices. The other explanation is that, as a result of choice experience, preferences or opportunity sets have changed (structural state dependence). It is well-known that this fundamental identification problem cannot be solved without imposing theoretical restrictions on the model.

We start by proposing a theoretical characterization of intertemporal choice models under pure taste persistence. Since models with pure taste persistence represent a reference case it is important to characterize this case theoretically as a point of departure for specifying models that allow for state dependence effects. Our characterization of choice behavior under pure taste persistent preferences can be viewed as an intertemporal version of Luce's axiom "Independence from Irrelevant Alternatives", (IIA). The IIA assumption states that the agent is (on average) only concerned with the alternatives presently in his choice set, i.e., alternatives outside his choice set are irrelevant.[1] The intertemporal version of IIA can briefly be described as follows: Consider the particular case in which the past choice sets are all the same but where the choice set in the current period is expanded to include new alternatives that were never feasible before. Under pure taste persistence the probability of choosing an alternative among the *new* alternatives that enter the choice set is independent from any choice in the past. Note that this does not immediately imply that the model has the Markov property, since we also must consider situations when the current choices also were feasible in the past. Under suitable regularity conditions we demonstrate that the intertemporal version of IIA implies a random utility representation where the utilities associated with each alternative are independent extremal processes. By drawing on results obtained by Dagsvik (1983 and 1988) it follows that when the choice sets are constant over time, the extremal utility processes yield a choice model which is a Markov chain (in continuous time) where the transition probabilities have a particular structure as a function of the choice set and the parameters of the utility processes. In Section 5 we demonstrate that the structure of this model is incompatible with the proportional hazard rate formulation. This result suggests that the proportional hazard rate formulation may be inappropriate for analyzing duration data which are generated from intertemporal choice behavior.

In many applications, it is of fundamental interest to allow choice sets to vary over time. For example, in labor market analyses it is important to be able to account for time varying sets of job opportunities and the fact that people are laid off. In Section 6 the model is extended to allow for particular cases of time varying choice sets. It turns out that the Markovian structure still holds as long as choice sets are not shrinking, but breaks down when alternatives are removed from the choice set. Unfortunately, only special cases are considered here; it is still an unsolved problem to derive the mathematical structure of the model in the general case.

In Section 7 we discuss how the framework can be extended to allow for structural state dependence effects. It turns out that the intertemporal version of IIA allows one to separate state dependence effects from spurious effects due to unobservables that affect preferences.[2]

An interesting question is whether the modelling framework developed here allows for an interpretation that is consistent with dynamic programming. It is well known that in the standard life cycle setting, the notion of two stage budgeting applies. This is also the case when some of the commodities are discrete, such as in labor supply models. (See for example, Blundell and Walker, 1986.) The notion of two stage budgeting means that the total expenditure in the current period is chosen in the first stage, while the within-period allocation of consumption is chosen in the second stage conditional on the current prices and total (chosen) expenditure. Thus the second stage decision is, theoretically, reduced to a static one (conditional on total expenditure and past choices). Unfortunately, the corresponding inference problem may in general be more complex. If unobservables affect utilities, then the period-specific total expenditure may be endogeneous. However, we are able to demonstrate that in the present setting where the utilities are extremal processes the chosen expenditure path can, in the second stage allocation, be treated as if it were exogenous. Thus, the present framework can be applied to specify a model for the second stage discrete choices, conditional on the chosen expenditure path. However, to estimate all the parameters of the model it is usually necessary also to solve a dynamic programming problem to determine the first stage allocation of expenditures.

The recent advances in dynamic programming, (see for example Rust (1994)), allows the researcher, in principle, to formulate and estimate quite general discrete choice models in a dynamic programming setting. Although the dynamic programming framework, as developed by Rust and others, is a very powerful methodology, it seems so far unable to handle models with preferences that are correlated over time, due to computational difficulties. However, a more fundamental problem from a theoretical point of view is the fact that the Bellman principle of optimality implies no testable restrictions, cf. Rust (1994), p.p. 3125-3130. Specifically, without additional theoretical restrictions the dynamic programming setting cannot resolve the type of identification problem discussed above. Thus, from a structural point of view predictions from dynamic programming models may be totally misleading, due for example to unidentified heterogeneity or state dependence effects.

Reduced form or myopic models choice

The present framework offers a structural alternative to the reduced form estimation implied by the proportional hazard rate framework. It offers new possibilities in this context that enables the analyst to attack previously unresolved identification problems.

The framework discussed in this paper also applies to analyze panel data on discrete choices. Maddala (1987), Fischer and Nijkamp (1987), Bolduc (1992), and Reader (1993) provide examples of such analyses. From a behavioral point of view some of these applications can be interpreted as analyses of myopic choice settings which are structural at each given point in time. As discussed above, they may, however, represent inappropriate formulations of myopic choice behavior over time if the purpose is to assess the significance of structural state dependence.

A particular myopic choice setting which is becoming increasingly popular is repeated or panel stated preference surveys. In stated preference surveys respondents are asked to express preferences for hypothetical products characterized by specific attributes. Such surveys enable the analyst to assess the potential demand for new products that are not available in the market. When each individual in a selected sample is presented with a series of stated preference experiments the problem of memory effect and taste persistence arises. Specifically, even if product attributes presented to the individuals vary across experiments, an individual's utility of a particular product may be correlated over time. The framework discussed in this paper explicitely models such taste persistence and has been applied by Dagsvik et al. (1996) to analyze the potential demand for alternative fuel vehicles, based on data from a stated preference survey.

Labor supply in a life cycle context

Consider a worker that faces the choice between the states "employed" and "not employed" in a life cycle context under uncertainty, cf. Heckman (1981b), Blundell and Walker (1986), and MaCurdy (1985). Provided the agent is boundedly rational in the sense that he does not take into account that current behavior may affect future distributions of income possibilities, we can formulate this as a two stage budgeting problem, as did Blundel and Walker (1986). Conditional on the savings decisions determined in the first stage we demonstrate in Section 4 that we can analyze the transitions into and out of employment as if the path of total expenditures were exogenous. The problem Heckman (1981b) was particularly concerned with was whether or not women's tastes for work are affected by work experience. The model extension presented in Section 7 allows one to identify structural state dependence effects of this type provided one is willing to accept the intertemporal version of IIA proposed here.

The paper is organized as follows: In Section 2 the choice setting is formally described. In Section 3 the main assumptions and results are discussed. In Section 4 we demonstrate that our framework allows for an interpretation that is consistent with optimizing behavior in a life cycle context where the (chosen) expenditure path can be treated as if it were exogenous when analyzing the within period allocations. In Section 5 we briefly discuss the implied structure of the choice process and interpretations in the context of econometric specification of hazard functions and transition intensities. In Section 6, we analyze the case with time varying choice sets, and in Section 7 we consider the extension of the framework to allow for state dependence.

2. The choice setting

The individual decision-maker (agent) is supposed to have preferences over a finite set of alternatives. Future preferences are assumed random (to the agent himself) in the sense that they vary from one moment in time to the next in a way that cannot fully be predicted by the agent. Alternatively, one may interpret the utilities as deterministic to the agent but random to the observer due to variables that are perfectly foreseeable to the agent but unobserved by the analyst.

Let S be the index set of m alternatives, a_1, a_2,..., a_m, and let $\Im$ be the index set that corresponds to the collection of all non-empty subsets from S. To each alternative, a_j, there is associated a stochastic process, $\{U_j(t), t \geq 0\}$, where $U_j(t)$ is the agent's (conditional indirect) utility of a_j given the information and choice history at time t. Moreover, $U_j(t) = v_j(t) + \varepsilon_j(t)$, where $v_j(t)$ is a deterministic component that may depend on alternative-specific attributes, and $\varepsilon_j(t)$ is a stochastic term. The agent chooses a_j at age t if $U_j(t)$ is the highest utility at t. Here age (time) is continuous. Let $\{J(t, B(t))\}$ denote the choice process, i.e.,

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

where $\{B(t), t > 0, B(t) \in \Im\}$ denotes the choice set process. We define the choice set process to be *increasing* at time t if $B(t) \setminus B(t-)$ is non-empty, and *decreasing* if $B(t-) \setminus B(t)$ is non-empty.[3] If $B(t) = B(s)$ for all s and t the choice set process is constant. Let $h(t) = \{J(s, B(s)), s < t\}$ denote the choice history and define $\mathbf{U}(t) = (U_1(t), U_2(t), \ldots, U_m(t))$, $\boldsymbol{\varepsilon}(t) = (\varepsilon_1(t), \varepsilon_2(t), \ldots, \varepsilon_m(t))$ and $\mathbf{v}_j(t) = (v_1(t), v_2(t), \ldots, v_m(t))$. We assume that $\{\mathbf{U}(t)\}$ is separable and continuous in probability.[4] Moreover, we assume that the cumulative distribution function (c.d.f.) of $\mathbf{U}(t)$ is absolutely continuous for any $t \in R_+$. This implies that there are no ties, that is

$$P\left(U_i(t) = U_j(t)\right) = 0.$$

3. Characterization of pure-taste-persistent preferences and choice probabilities

When the finite dimensional distributions of the utility process $\{\mathbf{U}(t), t > 0\}$ have been specified it is in principle possible to derive joint choice probabilities for a sequence of choices. However, since the class of intertemporal random utility models is quite large it is desirable to restrict this class on the basis of behavioral arguments. A related problem is that it seems to be rather difficult to find stochastic processes that are convenient candidates for utility processes in the sense that they imply tractable expressions for the choice probabilities in the intertemporal context.

In the present section we propose a behavioral assumption that enable us to characterize preferences and the choice probabilities in the "reference case", where there are no effects from past experiences on future preferences nor on future choice opportunities. In this reference case we say that the agents have *"pure-taste-persistent preferences"* (PTPP). Once we have obtained a theoretical characterization of the model in the case with pure taste persistence, then we may use this reference model as a point of departure for extending the model to allow for state dependence. The extension of the model to allow for state dependence will, however, be deferred to Section 7.

One way of introducing structural restrictions into the model is to apply probabilistic versions of the assumption of rational behavior. A famous example of this type of assumption is Luce Choice Axiom; "Independence from irrelevant alternatives", (IIA) (cf. Luce, 1959). A first attempt to extend IIA to the intertemporal setting was made by Dagsvik (1983).[5] Below we shall discuss the implications from another version of IIA, which is stated below.[6]

Assumption A1

Let $B(s)=B_1$ *, for all* $s<t$ *,* B_1 *Î Á and let* B_2 *Î Á be such that* $B_2 \setminus B_1 \neq \emptyset$. *Then for* $j \in B_2 \setminus B_1$ *,*

$$P\Big(J(t,B_2)=j \,\Big|\, J(s,B_1),'' \; s<t\Big)=P\big(J(t,B_2)=j\big). \tag{3.1}$$

It is important to stress that (3.1) does *not* mean that $\{J(t,B(t)),t>0\}$ is a Markov chain, nor is it a Bernoulli process.[7] This is so because (3.1) is assumed to hold only when $j \notin B_1$ and is silent about the relationship between the choices at different points in time when $j \in B_1$. Assumption A1 states that when $j \in B_2 \setminus B_1$ the event, "a_j is the preferred alternative in B_2", is stochastically independent of the preference orderings in B(s), for $s<t$. It is therefore natural to interpret Assumption A1 as an extension of the IIA property. The intuition is that even if previous choices provide information about the preferences over the alternatives in the "old" choice set, these choices provide no information about the utilities of the "new" alternatives, since they were not feasible in the past.

Remark 1

Assumption A1 can in fact be interpreted as our formalized notion of PTPP.

Assumption A2

At each point in time the distribution of the random term, ***e****(t) does not depend on* ***n****(t).*

Assumption A3

For any $t > 0, j \in S,$ *and any real number x there exists a value of* $v_j(t)$ *such that* $v_j(t) = x$.

Assumption A4

Apart from a location shift the finite dimensional probability laws of the indirect utility process, $\{max_k U_k(t), t \geq 0\}$*, are the same as the finite dimensional laws of* $\{U_1(t), t \geq 0\}$.

Assumption A2 states that at each moment in time the random term of the utility function is independent of the structural term. Assumption A3 states that the structural term of the utility function can vary over the whole real line when attributes vary freely.

Assumption A4 means that the utility processes are max-stable processes (see de Haan, 1984). The finite dimensional distributions of a max-stable process belong to the class of multivariate extreme value distributions.

Dagsvik (1995) has demonstrated that there is no loss of generality in assuming A4 since, in the absence of structural state dependence effects, any intertemporal random utility model can, under suitable regularity conditions, be approximated arbitrarily closely by choice probabilities generated from max-stable utilities.

Proposition 1

Assume that A2 and A3 hold. Then for any $B \,Î\, Á$, *Assumption A1 implies that*

$$P\big(J(t,B) = j\big) = \frac{e^{v_j(t)}}{\sum_{k \,Î\, B} e^{v_k(t)}} \tag{3.2}$$

where $v_j(t) = aEU_j(t)$ *and* $a > 0$ *is an arbitrary constant.*

Proof:

Recall that $\{\mathbf{U}(t), t > 0\}$ is continuous in probability. Recall also that since the utilities are independent of the choice set process, we are allowed to specify any sequence of choice sets which is useful for deriving implications about the preferences. To this end, let $B(t) = B$ and $B(t-) = B \setminus \{j\}$. Then A1 implies that

(3.3)

$$P\left(U_j(t) = \max_{k \in B} U_k(t), U_i(t-) = \max_{k \in B \setminus \{j\}} U_k(t-)\right) = P\left(U_j(t) = \max_{k \in B} U_k(t)\right) P\left(U_i(t) = \max_{k \in B \setminus \{j\}} U_k(t)\right).$$

By Theorem 50, p. 354, in Luce and Suppes (1965), (3.3) implies that (3.2) is a Luce model. Finally, Strauss (1979), p.p. 42-43, has demonstrated that the parameters $\{v_j(t)\}$ of the choice model are related to the utility function by $v_j(t) = \alpha E U_j(t)$, apart from an additive constant.

Q.E.D.

Remark 2

Without loss of generality we shall in the following put α=1. Note also that A4 is not needed to prove Theorem 1.

Let us now proceed by investigating the intertemporal structure of the random utilities that follows from A1.[8]

Above we postulated the existence of random utility processes such that A1 to A3 hold. It remains, however, to demonstrate that such processes really exist. In the one-period case Yellott (1977) and Strauss (1979) have, under different sets of conditions, demonstrated the equivalence between IIA and extreme value distributed utilities in a random utility model with independent utilities. We state a version of this result in the next theorem.

Proposition 2

Assume that A1 to A4 hold. Then the utility processes, $\{U_j(t), t \geq 0\}$, *j=1,2,...,m, are independent at each point in time and have type III extreme value distributed marginals.*[9]

Proof:

It follows from A4 that the joint distribution of $\left(U_1(t), U_2(t), \ldots, U_m(t)\right)$ belongs to the class of c.d.f. considered in Theorem 6 of Strauss (1979) with $\varphi(x) = e^{-x}$. The result now follows from Strauss, Theorem 6.[10]

Q.E.D.

Assumption A5

The utility processes $\left\{U_j(t), t \geq 0\right\}$, $j = 1,2,\ldots,m$, *are stochastically independent.*

Remark 3

Recall that two stochastic processes $\left\{U_i(t), t \geq 0\right\}$ and $\left\{U_j(t), t \geq 0\right\}$ may be stochastically dependent even if $U_i(t)$ and $U_j(t)$ are stochastically independent at each point in time. For example, $U_i(t)$ and $U_j(s)$ may be interdependent for $s \neq t$, even if $U_i(t)$ and $U_j(t)$ are independent. However, it seems plausible that in many applications the correlation between $U_i(t)$ and $U_j(s)$ is less than the correlation between $U_i(t)$ and $U_j(t)$, which implies that the utility processes are independent when the utilities at each point in time are independent.

The next result is the main result of this paper.

Theorem 1

Assume A1 to A5. Then the utilities are extremal processes with type III extreme value marginal distribution.

The proof of Theorem 1 is given in the appendix.

The class of extremal processes was introduced in statistics by Dwass (1964, 1966, 1974) and Tiago de Oliveira (1968, 1973). An extension to inhomogeneous extremal processes has been made by Weissman (1975). For our purpose it will be convenient to work with a modified inhomogeneous extremal process. The modified extremal process differs from the (standard) inhomogeneous extremal process by a deterministic time trend. More precisely, a modified inhomogeneous extremal utility function will be defined as processes $\left\{U_j(t), t > 0\right\}$, $j = 1,2,\ldots,m$, given by

$$U_j(t) = \max\left(U_j(s) - (t-s)\theta, W_j(s,t)\right), \tag{3.4}$$

$s < t$, where $U_j(0) = -\infty$, and where $W_j(s,t)$ is independent of $U_j(s)$ and has cumulative distribution function

$$P\left(W_j(s,t) \le y\right) = \exp\left(-\left(e^{v_j(t)} - e^{v_j(s)-(t-s)\theta}\right)e^{-y-\gamma}\right) \tag{3.5}$$

for $y \in R$, where $\gamma = 0.5772\ldots$ (Euler's constant), $\theta > 0$ is a constant and $\{v_j(t)\}$ are deterministic functions of t such that $v_j(t) + \theta t$ is nondecreasing for all j. Moreover, W(s,t) and W(s',t') are independent when $(s,t) \cap (s', t') = \varnothing$. It follows readily that $v_j(t)$ has the interpretation

$$v_j(t) = E\, U_j(t). \tag{3.6}$$

Tiago de Oliveira (1973) has demonstrated that when $v_j(t)$ is time constant then $U_j(t)$ becomes (strictly) stationary. As demonstrated by Resnick and Roy (1990) we can express a particularly version of the autocorrelation function of the utility process (3.4) as

$$\operatorname{corr}\left(\exp\left(-U_j(s)\right), \exp\left(-U_j(t)\right)\right) = \exp\left(v_j(s) - v_j(t) - (t-s)\theta\right). \tag{3.7}$$

Eq. (3.7) shows that when $v_j(t)$ varies slowly over time then the autocorrelation function is close to $\exp\left(-(t-s)\theta\right)$. In other words, the parameter θ characterizes the strength of time persistence in the preferences. Note that from a theoretical point of view it does not matter whether we use a modified extremal process or a (standard) extremal process since the time trend (θt) cancels in utility comparisons. However, the modified extremal process formulation allows a convenient interpretation due to (3.6) and (3.7).

Recall that $v_j(t) + \theta t$ must be nondecreasing for all j. This is inconvenient in practical empirical analysis. We shall therefore introduce a reparametrization given by

$$w_j(t) = v_j(t) + \log\left(\frac{\theta + v_j'(t)}{\theta}\right) \tag{3.8}$$

where we now have assumed that $v_j(t)$ is differentiable. From (3.8) it follows that

$$\exp\left(E\, U_j(t)\right) = \exp\left(v_j(t)\right) = \theta \int_0^t \exp\left(w_j(\tau) - (t-\tau)\theta\right) d\tau. \tag{3.9}$$

This particular reparametrization implies that $v_j(t) + \theta t$ is increasing for any $\{w_j(t), t > 0\}$. However, the main motivation behind (3.8) is that the reparametrization above is interesting for theoretical reasons. To realize this note first that when Δt is small we get from (3.5) and (3.9) that

$$P\left(W_j\left(t-\Delta t, t\right)\leq y\right)=\exp\left(-\theta\Delta t\, e^{w_j(t)-y-\gamma}+o(\Delta t)\right) \tag{3.10}$$

which shows that $w_j(t)$ has, apart form an additive term, the interpretation as the mean of "instantaneous" utility increments, $\left\{W_j\left(t-\Delta t, t\right)\right\}$. When $w_j(t)$ is independent of time (3.9) reduces to

$$\exp\left(E\,U_j(t)\right)=e^{w_j}\left(1-e^{-\theta t}\right). \tag{3.11}$$

Thus for large t, a constant mean utility level corresponds to constant $\left\{w_j(t)\right\}$. Also from (3.9) we realize that θ is analogous to a rate of preference parameter because by (3.9), the mean utility at time t can be expressed as an integral of past weighted "instantaneous" mean utilities. Specifically, the contribution from the period s-specific systematic utility component to the current mean utility is evaluated by multiplying $\exp\left(w_j(s)\right)ds$ by the "depreciation" factor $\exp\left(-(t-s)\theta\right)$. [11] This depreciation factor accounts for the loss of memory and/or decrease in taste persistence as the time lag increases.

To clarify the interpretation further, consider the autocorrelation function (3.7) with constant $\left\{w_j(t)\right\}$. Then (3.7) reduces to

$$\operatorname{corr}\left(\exp\left(-EU_j(s)\right),\exp\left(-EU_j(t)\right)\right)=\frac{1-e^{-\theta s}}{1-e^{-\theta t}}\cdot e^{-(t-s)\theta}. \tag{3.12}$$

Thus when s and t are large the mean utility in this case equals w_j, (apart from an additive constant) and the autocorrelation function becomes exponential.

Remark 4

It is important to emphasize that in the discussion of the extremal process above we have made no assumptions that restricts the class of inhomogeneous extremal processes with extreme value marginals.

In the following we shall use the concept of *modified extremal process,* to mean a stochastic process which satisfies (3.4) and (3.5) with $U_j(0)=-\infty$, and with $v_j(t)$ differentiable in t for all j.

Theorem 2

Assume that the random utilities are independent modified extremal processes. Assume that the choice set process is constant ocer time. Then the indirect utility, $max_{k \in B} U_k(t)$, *is independent of* $\{J(t, B), t \leq t\}$ *for any* $B \in \Im$. *Moreover, the choice process* $\{J(t, B), t > 0\}$ *is a Markov chain.*

A proof of this result has been given by Resnick and Roy (1990). The fact that $\{J(t, B), t > 0\}$ is a Markov chain was originally proved by Dagsvik (1983).

Dagsvik (1988) and Resnick and Roy (1990) extend the result of Theorem 4 to the case where $\{\mathbf{U}(t), t \geq 0\}$ is a multivariate extremal process. Dagsvik considers the case where **U**(t) ~~at~~ each t ~~has~~ a type III multivariate extreme value distribution that is absolutely continuous. The resulting (marginal) choice probabilities at a given point in time in this case become generalized extreme value probabilities. Resnick and Roy (1990) allow **U**(t) to have a multivariate c.d.f. that is not necessarily absolutely continuous.

Recall that by (3.4) the utility processes are Markov processes. However, utility processes with the Markov property do not usually imply that the corresponding choice process {J(t,B)} is Markovian. For example, Gaussian utility processes with the Markovian property do not imply that the choice process is Markovian. In fact, there exist no Gaussian utility processes in continuous time that can generate Markovian choice models.

4. Life cycle consistent choice behavior

The purpose of this section is to demonstrate that by a suitable extension of the model it follows that the results obtained above are consistent with choice behavior in a life cycle context. We consider the following setting: The agent's problem is to choose the level of consumption of an infinitely divisible composite good and a discrete alternative (state) from a set of mutually exclusive alternatives in each period (time is discrete). Let α_{jt} be the period specific cost of choosing the discrete alternative a_j, and c_t the consumption in period t with price index p_t. There are no transaction costs and the expectation of future uncertain events are not affected by current and past behavior. Let y_t denote total expenditure in period t and let r_t be the interest rate in period t. Furthermore, let ω_t be the income in period t and Y_t the wealth at the beginning of period t. The horizon is assumed finite. To the agent, future preferences are assumed known with perfect certainty, so the random components of the utilities are entirely due to the econometrician's lack of information about variables that affect the decision-making process. Future income, prices, interest rates and costs are uncertain. For simplicity we assume that incomes, interests, expenditures and wage rates are discrete variables (for example, with NOK as unit).

Let $J^*(t, y_t)$ denote the index of the chosen alternative in period t given total expenditure y_t in period t. The budget constraints in period t are given by

$$y_t = Y_t + \omega_t - \frac{Y_{t+1}}{1+r_{t+1}} - T(\omega_t, Y_t) \tag{4.1}$$

and

$$y_t = \sum_i \alpha_{it} 1_{\{J^*(t,y_t)=i\}} + c_t p_t \tag{4.2}$$

where $T(\cdot)$ is the tax function.[12] Let $U_j(t,c)$ denote the instantaneous utility of (a_j, c), as of period t. We assume that $\mathbf{U}(t) \equiv (U_1(t,1), U_2(t,1), \ldots, U_m(t,1), U_1(t,2), \ldots)$ is a multivariate modified extremal process with joint distribution in period t equal to

$$P(\mathbf{U}(t) \le \mathbf{u}) = \exp\left\{-\left[\sum_{c\ge 1}\left(\sum_{j=1}^{m} \exp(v_j(t,c) - u_j(c))\right)^{1/\sigma}\right]^{\sigma}\right\} \tag{4.3}$$

where $v_j(t,c) = E\,U_j(t,c) - 0.5772$, and $0 < \sigma \le 1$ is a parameter that has the interpretation

$$\text{corr}(U_j(t,c), U_j(t,c')) = 1 - \sigma^2 \tag{4.4}$$

for $c \ne c'$. Moreover, (4.3) implies that for $j \ne k$, and all c and c',

$$\text{corr}(U_j(t,c), U_k(t,c')) = 0. \tag{4.5}$$

(For a general description of multivariate extremal processes, see Dagsvik, 1988.) Thus, the utility structure (4.3) is compatible with a particular nested generalized extreme value structure where the utilities are independent, given a particular consumption level, but where the utilities of (a_j, c) and (a_j, c') are correlated. Also it follows that for a given c, $\{U_j(t,c), t > 0\}$, $j = 1,2,\ldots,m$, are independent modified extremal processes (cf. McFadden, 1981, p.p. 228-230, for a more detailed discussion of this type of GEV models). Let $V_j(t, Y_t)$ denote the value function as of period t given that $J^*(t, y_t) = j$ and given Y_t, i.e., $V_j(t, Y_t)$ is the highest expected utility attainable, conditional on current wealth Y_t and current choice of the discrete alternative. Under the assumption of additive intertemporal separability, the Bellman equation that corresponds to dynamic choice behavior under uncertainty is given by

$$V_j(t,Y_t)=U_j\left(t,y_t-\alpha_{jt}\right)+\rho E_t\left(\max_k V_k\left(t+1,Y_{t+1}\right)\Big| J^*(t,y_t)=j\right) \tag{4.6}$$

where $\rho<1$, is the time-preference discounting factor and E_t denotes the subjective expectation operator given the agent's information at time t.

As is well-known, the decision problem above can be viewed as a two stage process in which the agent determines the expenditure path in the first stage and the choice of state in each period is determined conditional on the expenditure path. In particular, the second stage allocation is a pure static choice problem that is solved by maximizing $U_j\left(t,y_t-\alpha_{jt}\right)$ when total expenditure, y_t, has been determined. As is also well-known, the application of two stage budgeting to undertake life cycle consistent empirical analyses may be difficult in the presence of unobservables because the optimal expenditure path may in general be correlated with the random terms of $\left\{U_j\left(t,y_t-\alpha_{jt}\right), j=1,2,\dots,m, t=1,2,\dots\right\}$. However, due to the properties of the extremal processes we shall see in a moment that this will not be the case here.

Let

$$\tilde{U}(t,y)\equiv\max_k U_k\left(t,y-\alpha_{kt}\right)$$

and

$$V(t,Y)\equiv\max_k V_k(t,Y).$$

Since we assume that preferences are exogenous it follows from (4.6) that y_t is determined by

$$y_t=\arg\max_y\left(\tilde{U}(t,y)+\rho E_t V\left(t+1,\left(1+r_{t+1}\right)\left(Y_t+\omega_t-T\left(\omega_t,Y_t\right)-y\right)\right)\right) \tag{4.7}$$

and

$$V\left(t,Y_t\right)=\tilde{U}\left(t,y_t\right)+\rho E_t V\left(t+1,Y_{t+1}\right). \tag{4.8}$$

From (4.3) and the assumption that **U**(t) is a multivariate modified extremal process it follows that

$$\tilde{\mathbf{U}}(t)=\left(\tilde{U}(t,1),\tilde{U}(t,2),\dots\right)$$

is a modified multivariate extremal process. Recall that the extremal process has the markov property. The term

$$E_t\, V\Big(t+1, \big(1+r_{t+1}\big)\big(Y_t+\omega_t - T\left(\omega_t, Y_t\right)-y\big)\Big)$$

will therefore depend on $\tilde{\mathbf{U}}(t)$ and random terms that are associated with future preferences and therefore are independent of **U**(t). Consequently, we can write

$$y_t = f_t\left(\boldsymbol{\eta}(t), \boldsymbol{\xi}(t)\right)$$

for some function $f_t(\cdot)$, where $\boldsymbol{\eta}(t)$ denote the random elements of $\tilde{\mathbf{U}}(t)$ and $\boldsymbol{\xi}(t)$ is a vector of random variables that are independent of $\eta(t)$ and also independent of **U**(t). From Lemma 2, which is proved in the appendix, it follows that $\boldsymbol{\eta}(t)$ and $J^*(t,y)$ are stochastically independent for each given y. But then it follows that

$$P\left(J^*\left(t, y_t\right)=j \,\middle|\, y_t = y\right) = P\left(J^*(t,y)=j \,\middle|\, f_t\left(\eta(t), \xi(t)\right)=y\right) = P\left(J^*(t,y)=j\right).$$

We can thus analyze the choice process $\left\{J^*\left(t, y_t\right), t>0\right\}$, conditional on $\left\{y_t, t>0\right\}$, as if $\left\{y_t, t>0\right\}$ were exogenous. From the results in Section 7 it follows that the results in this section continue to hold in the presence of state dependence, as long as the anticipation of future uncertain events is not affected by current and past choices.

5. Some implications for econometric specifications of transition intensities under pure taste persistence

The results obtained above are useful for justifying the choice of functional form of the likelihood function of observations on $\left\{J(\tau, B),\ \tau \le t\right\}$ for a particular agent under PTPP. The first step in specifying an empirical model is to specify the structural parts of the model.

Recall that according to (3.10) it is possible to express the structural term of the current utility as a "depreciated" sum of the structural parts of the past increments. This allows us to interpret $w_j(t)$, or equivalently $\exp\big(w_j(t)\big)$, as the representative instantaneous utility of alternative a_j at time t. In empirical applications one would typically specify $w_j(t)$ as

$$w_j(t) = w\big(\mathbf{X}_j(t)\big) \tag{5.1}$$

where $\mathbf{X}_j(t)$ is a vector of observable attributes specific to alternative a_j and $w(\cdot)$ is a suitably chosen functional form that is known apart from an unknown vector of parameters. If we apply the results in Dagsvik (1988) we get the next result.

Theorem 3

Assume that the utilities are independent modified extremal processes and the choice sets process is constant over time. Then for B Î Á

$$P\left(J(t,B)=j\right)=\frac{\int_0^t e^{w_j(t)-(t-t)q}\,dt}{\sum_{k\in B}\int_0^t e^{w_k(t)-(t-t)q}\,dt}, \tag{5.2}$$

and

$$P\left(J(t,B)=j \mid J(s,B)=i\right)=\frac{\int_s^t e^{w_j(t)-(t-t)q}\,dt}{\sum_{k\in B}\int_0^t e^{w_k(t)-(t-t)q}\,dt}, \tag{5.3}$$

for $i \neq j$ *and* $P\left(J(t,B)=i \mid J(s,B)=i\right)$ *is determined by the adding-up condition. Moreover*

$$Corr\left(exp\left(-\max_{k\in B}U_k(s)\right), exp\left(-\max_{k\in B}U_k(t)\right)\right)=\frac{\sum_{k\in B}\int_0^s e^{w_k(t)-(s-t)q}\,dt}{\sum_{k\in B}\int_0^t e^{w_k(t)-(t-t)q}\,dt}\,e^{-(t-s)q}\,. \tag{5.4}$$

Proof:

The results (5.2) and (5.3) follow from Dagsvik (1988) by inserting (3.9). Eq. (5.4) follows from Resnick and Roy (1990).[13]

Q.E.D.

Before we proceed we shall recall the formal definition of transition intensities. We give the definition in the general case where the choice process may depend on past choices. This is of relevance for the discussion in Sections 6 and 7. Provided the choice sets do not change over time the transition intensities of $\{J(t,B), t>0\}$ are given by

$$q_{ij}\left(t, h(t)\right) \equiv \lim_{s \to t} \frac{P\left(J(t,B)=j \mid J(s,B)=i, h(s)\right)}{t-s} \tag{5.5}$$

for $i \neq j$, and

$$q_{ii}\left(t, h(t)\right) \equiv \lim_{s \to t} \frac{\left(P\left(J(t,B)=i \mid J(s,B)=i, h(s)\right)-1\right)}{t-s}. \tag{5.6}$$

The transition probabilities given that a transition occurs are defined by

$$\pi_{ij}\left(t, h(t)\right) \equiv \lim_{s \to t} \frac{P\left(J(t,B)=j \mid J(s,B)=i, h(s)\right)}{\sum_{k \in B} P\left(J(t,B)=k \mid J(s,B)=i, h(s)\right)} = \frac{q_{ij}\left(t, h(t)\right)}{q_{ii}\left(t, h(t)\right)}. \tag{5.7}$$

The interpretation of (5.7) is as the transition probability of going to alternative a_j at time t given that alternative a_i is left and given the choice history prior to t. The next result is immediate.

Corollary 1

Under the assumptions of Theorem 5 it follows that

$$q_{ij}\left(t, h(t)\right) = q_{ij}(t) = \frac{e^{w_j(t)}}{\sum_{k \in B} \int_0^t e^{w_k(t)-(t-t)q} \, dt}, \tag{5.8}$$

$$p_{ij}\left(t, h(t)\right) = p_{ij}(t) = \frac{e^{w_j(t)}}{\sum_{k \in B \setminus \{i\}} e^{w_k(t)}} \tag{5.9}$$

for $i \neq j$, and

$$q_{ii}\left(t, h(t)\right) = q_{ii}(t) = -\sum_{k \in B \setminus \{i\}} q_{ik}(t). \tag{5.10}$$

Let us next consider the particular case where $\left\{w_j(t)\right\}$, $j = 1,2,\ldots,m$, are constant over time i.e., $w_j(t) = w_j$. Then (5.8) reduces to

$$q_{ij}(t) = \frac{\theta e^{w_j}}{\left(1-e^{-t\theta}\right)\sum_{k \in B} e^{w_k}} \equiv \frac{\theta P_j}{1-e^{-\theta t}}, \tag{5.11}$$

for $i \neq j$ where P_j is the probability of being in state j. Recall that by (3.7) and (3.12) the degree of taste persistence in the indirect utility can be measured by θ. Specifically, when θ is large there is little taste persistence (provided t is large) while when θ is close to zero tastes are strongly correlated over time. Moreover, (5.11) shows that the transition intensities are stationary when t is large. However, when t is small then the transition intensities given by (5.11) depend on time. This is due to the fact that in the beginning of a choice process the length of the choice history (age) will influence the strength of the taste persistence effect.

Observe that the structure of (5.11) can be viewed as a special case of the model in Olsen et al. (1986). However, as the utilities in their model are not serially correlated, θ in their model seems at first glance to yield a different interpretation. In their framework the utilities are viewed as independent draws that occur according to a Poisson process with intensity θ. But this means in fact that also in their setting θ allows the interpretation as a measure of taste persistence because when θ is small the random draws occur rarely and therefore preferences are rather stable over time. In contrast, when θ is large preferences are likely to change frequently. There is, however, an important and testable difference between the model in Olsen et al. and the present framework. While θ in their model may depend on the state occupied, the assumptions above imply that θ cannot depend on the state occupied in our model.

Let us finally compare the structure of the hazard rate with the proportional hazard rate framework which has been used extensively in empirical analyses, cf. Heckman and Singer (1985). Unfortunately, in this literature the motivation for the proportional hazard rate framework is totally ad hoc. To illustrate this, consider the following labor market example: A female worker has the choice between the two states "employed" (state 1) and "not employed" (state 2). For simplicity we assume a stationary setting, i.e., $w_j(\tau) = w_j$ where w_1 is a function of the wage rate the agent faces in the labor market, while w_2 depends on non-labor income, age and the number of small children in the household. When t is large it follows from (5.10) and (5.11) that the hazard rates have the structure

$$-q_{ii} = \theta\left(1 - \frac{e^{w_i}}{e^{w_1} + e^{w_2}}\right), \tag{5.12}$$

for $i = 1,2$. In contrast, within the proportional hazard rate framework one would typically specify q_{ii} as

$$-q_{ii} = \exp\left(\beta_{0i} + X\beta_i\right) \tag{5.13}$$

where X is a vector consisting of explanatory variables and (β_{0i}, β_i) is a vector of state-specific parameters. Thus the functional form in (5.12) that follows from utility maximizing behavior is different from the specification in (5.13). At best (5.13) can be interpreted as a reduced form model. In (5.12) the key parameters w_1, w_2 and θ have a clear interpretation, and θ and $w_1 - w_2$ can also be separately identified. This point is even more dramatic in the multistate case: In this case some authors (see for example Andersen et al. (1991)) have specified transition intensities $\{\tilde{q}_{ij}(t)\}$ on the form

$$\tilde{q}_{ij}(t) = \lambda_{ij}(t)\exp\left(f\left(X_i, X_j; b\right)\right) \tag{5.14}$$

where $f(\cdot)$ is some specified function, X_i is a time invariant vector of covariates that characterize alternative a_i, and b is a vector of parameters. Let us now compare the structure (5.14) with (5.8). We realize that (5.8) is essentially different from (5.14) in that (5.8) depends on all the covariates while (5.14) only depends on the covariates related to alternatives a_i and a_j. Also these covariates enter the model in a particular way. Finally, (5.8) depends on the choice set in a particular way. Therefore, the standard proportional hazard specification (5.14), which is often applied in duration analysis, is inconsistent with the present random utility formulation when the number of states is larger than two.

6. Allowing for time varying choice sets

In many applications it is of interest to allow for time varying choice sets. For example, in the analysis of labor market dynamics, workers' choice sets may change over time and result in periods of unemployment. We shall now discuss the extension of the model to allow for time varying choice sets.

Unfortunately, we are presently unable to provide an exhaustive discussion of the model in this case due to the complexity of the problem. For example, when the choice sets decrease over time the corresponding choice model will in general not have the Markov property. Below we shall therefore only discuss a few particular cases.

Let us start by considering the choice model at two points in time. The choice model in this case was obtained by Dagsvik (1983). For the sake of completeness and with the particular representation of the modified extremal process introduced in Section 3 we give the result below.

Theorem 4

Assume that the utilities are modified extremal processes. Then for $s < t$ *, we have*

$$P\big(J(s,B(s))=i,\, J(t,B(t))=j\big)$$
$$= \frac{P\big(J(s,B(s))=i\big)\int_0^t e^{w_j(t)-(t-t)q}\,dt - d_j(B(s))\int_0^s e^{w_j(t)-(t-t)q}\,dt}{\sum_{k\in B(t)}\int_0^t e^{w_k(t)-(t-t)q}\,dt + \sum_{k\in B(s)\setminus B(t)}\int_0^s e^{w_k(t)-(t-t)q}\,dt}$$
$$+ \frac{P\big(J(t,B(t))=j\big)\big(1-d_i(B(s))\big)\int_0^s e^{w_i(t)-(t-t)q}\,dt}{\sum_{k\in B(t)}\int_0^t e^{w_k(t)-(t-t)q}\,dt + \sum_{k\in B(s)\setminus B(t)}\int_0^s e^{w_k(t)-(t-t)q}\,dt} \tag{6.1}$$

where $d_i(B)=1$ *if* $i\in B$*, and zero otherwise, and*

$$P\big(J(t,B(t))=j\big)=\frac{\int_0^t e^{w_j(t)-(t-t)q}\,dt}{\sum_{k\in B(t)}\int_0^t e^{w_k(t)-(t-t)q}\,dt}. \tag{6.2}$$

Proof:

The results follow from (3.9) and Dagsvik (1983), pp. 32-33.

Q.E.D.

Theorem 5

Assume that the utilities are modified extremal processes. Assume moreover that the choice set process is nondecreasing. Then the choice process $\{J(t,B(t)),t>0\}$ *is a Markov chain. The transition probabilities are given by*

$$P\big(J(t,B(t))=j \mid J(s,B(s))=i\big)=P\big(J(t,B(t))=j\big)-z\big(s,t,B(s),B(t)\big)\,P\big(J(s,B(s))=i\big)d_i\big(B(t)\big) \tag{6.3}$$

for $i\neq j,\; i\in B(t),\, j\in B(t),$ *where*

$$z\big(s,t,B(s),B(t)\big)=\frac{\sum_{k\in B(s)}\int_0^s e^{w_k(t)-(t-t)q}\,dt}{\sum_{k\in B(t)}\int_0^t e^{w_k(t)-(t-t)q}\,dt}\cdot e^{-(t-s)q}\;, \tag{6.4}$$

and $d_i(B(t)) = 1$ *if* $i \in B(t)$, *and zero otherwise, and the state probabilities are given by (6.2).*

Proof:

Resnick and Roy (1990), have proved that the choice process is Markovian in this case. The structure of (6.3) follows from (6.1).

Q.E.D.

Remark 5

Similarly to (5.4) it can be demonstrated that $\zeta(s, t, B(s), B(t))$ can be interpreted as a measure of autocorrelation.

7. Extending the model to allow for state dependence

So far we have only discussed the functional form of the choice probabilities of $\{J(t, B(t))\}$ under PTPP. The question now arises how the particular functional form that follows from PTPP should be modified in the presence of state dependence. Notice first that when the utility processes are altered by the choice history a simultaneous equation bias problem arises. This is so because the structural terms of the utility processes become dependent on past choices, and consequently they will depend on past realizations of the utility processes.

For simplicity we shall consider the discrete time case. Accordingly, we will assume that the utility processes are independent modified experience-dependent extremal processes defined by

$$U_j(t) = \max\left(U_j(t-1) - \theta, W_j(t, h(t))\right) \tag{7.1}$$

where $W_j(t, h(t))$ is a random variable with distribution

$$P\left(W_j(t, h(t)) \le w \mid \mathbf{U}(t-1)\right) = \exp\left(-\exp\left(g_j(t, h(t)) - w\right)\right) \tag{7.2}$$

and where $g_j(t, h(t))$ is a parametric function of the attributes of alternative j and past choice experience. Define $v_j(t, h(t))$ recursively by

$$\exp\left(v_j(t, h(t)) + t\theta\right) = \exp\left(v_j(t-1, h(t-1)) + (t-1)\theta\right) + \exp\left(g_j(t, h(t))\right). \tag{7.3}$$

Note that when g_j does not depend on h(t) then $v_j(t,h(t))$ reduces to $v_j(t)=EU_j(t)$, and (7.1) reduces to (3.4).

The following result extends Theorem 5 to the case with state dependence.

Theorem 6

Assume that the choice set process is non-decreasing and $\{U_j(t), t \geq 0\}$, $j = 1,2,\ldots,m,$ *are independent and experience-dependent utility processes defined by (7.1) to (7.3). Then the (one step) transition probabilities, conditional on the choice history, are given by*

$$P\Big(J\big(t,B(t)\big)=j \,\Big|\, J\big(t-1,B(t-1)\big)=i,h(t)\Big)= R_j\big(t,h(t)\big)\Big[1-exp\Big(v_j\big(t-1,h(t-1)\big)-v_j\big(t,h(t)\big)-q\Big)\Big]$$
$$\equiv R_j\big(t,h(t)\big)-z\big(t-1,t,B(t-1),B(t),h(t)\big)\,R_j\big(t-1,h(t-1)\big)\,d_i\big(B(t)\big) \tag{7.4}$$

for $i \neq j,\ \ j \in B(t), i \in B(t-1),$ *and*

$$P\Big(J\big(t,B(t)\big)=i \,\Big|\, J\big(t-1,B(t-1)\big)=i,h(t)\Big)$$
$$= R_i\big(t,h(t)\big)+z\big(t-1,t,B(t-1),B(t),h(t)\big)\big(1-R_i\big(t-1,h(t-1)\big)\big)d_i\big(B(t)\big) \tag{7.5}$$

where

$$R_j\big(t,h(t)\big)=\frac{e^{v_j(t,h(t))}}{\sum_{k\in B(t)} e^{v_k(t,h(t))}}, \tag{7.6}$$

$$z\big(t-1,t,B(t-1),B(t),h(t)\big)=\frac{\sum_{k\in B(t-1)} e^{v_k(t-1,h(t-1)\ -q)}}{\sum_{k\in B(t)} e^{v_k(t,h(t))}}, \tag{7.7}$$

and $v_j\big(t,h(t)\big)$ *is defined by (7.3)*

A proof of Theorem 6 is given in the appendix.

Corollary 2

Under the assumptions of Theorem 6 the c.d.f. of the indirect utility, $max_{k\in B(t)} U_k(t)$*, depends on* $\{J(t, B(t)), t \le t\}$ *solely through*

$$\sum_{k\in B(t)} exp\left(v_k(t, h(t))\right).$$

Proof:

The result of Corollary 2 follows directly from Theorem 2, (7.1) and (7.2).

Q.E.D.

Corollary 2 implies that the life cycle consistent property discussed in Section 4 also holds in the case with utilities that are experience-dependent extremal processes, provided the agent does not take into account that current behavior may alter future preferences and future choice constraints.

It is important to notice that in contrast to (6.2), $R_j(t, h(t))$ can of course not be interpreted as the marginal choice probability at time t since it depends on the choice history. It can, however, be interpreted as the conditional choice probability at time t for an agent equipped with preferences that have been altered by experience.

We may, analogous to Section 5 model state dependence effects in the reparameterized version in which $\{v_j(t, h(t))\}$ is substituted by $\{w_j(t, h(t))\}$ defined by

$$e^{w_j(t,h(t))} = e^{v_j(t,h(t))} - e^{v_j(t-1,t,h(t-1))-\theta} \tag{7.8}$$

which implies that

$$e^{v_j(t,h(t))} = \sum_{\tau=1}^{t} e^{w_j(\tau,h(\tau))+(\tau-t)\theta}. \tag{7.9}$$

For notational simplicity, let $Q_{ij}(t-1, t, h(t))$ denote the one step transition probability of going from i at time t-1 to j at time t, given the history h(t). From (7.9), (7.4) and (7.6) it follows that $Q_{ij}(t-1, t, h(t))$ can be expressed as

$$Q_{ij}(t-1, t, h(t)) = \frac{e^{w_j(t,h(t))}}{\sum_{k\in B(t)} \sum_{\tau=1}^{t} e^{w_k(\tau,h(\tau))-(t-\tau)\theta}} \tag{7.10}$$

for $i \neq j$. The transition probability given a transition has a structure that is completely analogous to (5.9), i.e.,

$$\pi_{ij}(t-1,t,h(t)) = \frac{e^{w_j(t,h(t))}}{\sum_{k \in B(t)\setminus\{i\}} e^{w_k(t,h(t))}}. \tag{7.11}$$

We realize now that both $\{w_j(t,h(t))\}$ as well as the taste persistent parameter θ are separately identified. From (7.10) we get that

$$\log\left[\frac{Q_{ij}(t-1,t,h(t))}{Q_{21}(t-1,t,h(t))}\right] = w_j(t,h(t)) - w_1(t,h(t)). \tag{7.12}$$

Eq. (7.12) implies that $w_j(t,h(t)) - w_1(t,h(t))$ is non-parametrically identified. For example, if $w_j(t,h(t))$ has the structure

$$w_j(t,h(t)) = \sum_{r=1}^{s} \left(\beta_{1r}\, f_{1r}\left(X_j(t)\right) + \beta_{2r}\, f_{2r}\left(X_j(t),h(t)\right)\right) \tag{7.13}$$

where $\{f_{kr}\}$ are known functions and $\{\beta_{kr}\}$ are unknown parameters, $k = 1,2,\ r = 1,2,\ldots,s,$ then $\{\beta_{kr}\}$ are identified under rather general conditions on $\{f_{kr}(\cdot)\}$.

Finally, when $w_j(t,h(t))$ has been detemined, θ is identified because (7.10) implies that

$$\frac{e^{w_j(t,h(t))}}{Q_{ij}(t-1,t,h(t))} - \sum_{k \in B(t)} e^{w_k(t,h(t))} = e^{-\theta} \cdot \frac{e^{w_j(t-1,h(t-1))}}{Q_{ij}(t-2,t-1,h(t-1))} \tag{7.14}$$

for $i \neq j$.

Example (Heckman, 1981b)

Consider the labor supply example analyzed by Heckman (1981b). Let $U_2(t)$ be the utility of working and $U_1(t)$ the utility of not working. If we assume that the transition probabilities given by (7.10) are specified as

$$w_1(t,h(t)) = w_1(t) = X_1(t)\alpha_1 \tag{7.15}$$

and

$$w_2(t, h(t)) = X_2(t)\alpha_2 + \delta D(t-1), \tag{7.16}$$

where $X_1(t)$ is a vector that may consist of age, length of schooling and number of small children, $X_2(t)$ may be some function of the marginal wage rate (or instruments for the marginal wage rate), $D(\tau)$ is equal to one if the agent has worked in period τ and zero otherwise, α_1, α_2 and δ are parameters to be estimated. In the formulation above $\delta = 0$ implies PTPP, otherwise there is state dependence in that the agent's utility for work is affected by work experience.

Clearly, this model is identified and the specification (7.15) and (7.16) can be exploited to form the likelihood for a sample of individual work histories to estimate α_1, α_2, the taste persistence parameter θ and the state dependence parameter δ.[14]

8. Conclusions

In this paper we have considered the problem of obtaining a theoretical foundation for the choice of functional form and stochastic structure in intertemporal discrete choice models.

It is demonstrated that a particular extension of Luce IIA axiom implies a random utility model where the utilities are extremal processes. When the choice set process is non-decreasing this model has the Markov property with a particular structure of the transition probabilities. It is also demonstrated that this model is, under specific assumptions, consistent with optimizing behavior in a life cycle context where the (chosen) expenditure path can be treated as if it were an exogenous process in the probabilities that correspond to the within-period discrete choices.

Finally, we discuss how the choice model can be extended to allow for increasing and decreasing choice sets and structural state dependence. In the case with time varying choice sets the choice model does not in general have the Markov property. At present we are only able to derive the choice probabilities in a few particular cases. A general treatment of the case with time varying choice sets is left for future research.

The framework developed in this paper is analytically tractable and it therefore appears convenient for empirical applications.